Using household surveys for deriving labour market, poverty and inequality trends in South Africa

by

Kwan Cheung Derek Yu

Dissertation presented for the degree of Doctor of Economics in the Faculty of Economic and Management Sciences at the University of Stellenbosch

Promoter: Prof. Servaas van der Berg
Faculty of Economic and Management Sciences
Department of Economics

December 2012

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2012

ABSTRACT

In order to evaluate the extent to which South Africa achieve the objectives of poverty and inequality reduction as well as job creation, up-to-date and reliable data are required. Since the transition, various survey data have been commonly used for these analyses, namely Census, Community Survey (CS) 2007, Income and Expenditure Survey (IES), October Household Survey (OHS), Labour Force Survey (LFS), Quarterly Labour Force Survey (QLFS), General Household Survey (GHS), Project for Statistics on Living Standards and Development (PSLSD), National Income Dynamics Study (NIDS) and All Media Products Survey (AMPS).

However, these datasets are not fully comparable, due to differences in the sampling design, sample size, questionnaire structure, methodology to derive labour market status, as well as the way the income and expenditure information was collected. Hence, this dissertation begins by analysing these issues in each survey in Chapter 2. With regard to the income and expenditure information, it was collected differently in the surveys: the recall method was used in all surveys except IES 2005/2006, the only survey that adopted the diary method; respondents were asked to report the actual amount in some surveys but only asked to declare the relevant interval in others; for the former approach, respondents could either declare the single estimate amount or amounts for sub-categories that were then aggregated; for interval data, various methods can be used to determine the amount in each interval. Thus, Chapter 3 begins by discussing the merits and drawbacks of these approaches, as well as how they would affect the reliability and comparability of income and expenditure variables across the surveys.

In some surveys (e.g., the two censuses and CS 2007), quite high proportions of households incorrectly reported zero income or expenditure or did not specify their income or expenditure. Poverty and inequality estimates could be influenced by either including or excluding these households from the analyses. Hence, various approaches to deal with these households are examined in Chapter 3. As the surveys typically under-captured income or expenditure when compared with the national accounts income, the validity of the resultant poverty and inequality estimates might be affected. Hence, arguments for and against adjusting the survey means in line with the national accounts mean (e.g. by shifting the survey distribution rightwards) are discussed. As the survey data are, strictly speaking, cross-sectional and not designed for time-series labour market, poverty and inequality analyses, it is

sometimes argued that the data should be re-weighted to be consistent with demographic and geographic numbers presented by the Actuarial Society of South Africa (ASSA) and Census data. This cross entropy re-weighting approach is discussed in Chapter 3. Finally, the chapter examines the labour market status derivation methodology in all OHSs, LFSs and QLFSs in greater detail, and investigates how the changes across the surveys could possibly affect the comparability of labour market estimates throughout the years.

The dissertation then examines the labour market trends since the transition by using the OHS, LFS and QLFS data, and it is found that both the labour force and employment numbers increased in general since the transition, but the latter increase was not rapid enough to absorb the expanding labour force. In addition, the number of narrow unemployed doubled between 1994 and 2009, and the narrow unemployment rate showed an upward trend and peaked at just above 30% in 2003. It decreased between 2004 and 2007, before rising again in 2008-2009 due to the impact of global recession. Application of the cross entropy approach does not substantially affect labour market trends, suggesting that the trends (including the abrupt increase in labour market estimates during the changeover from OHS to LFS) were either real or took place due to the improvement of the questionnaire to capture the labour market status of the respondents better. Furthermore, the application of the LFS 2000b-LFS 2007b methodology on the earlier surveys reduced the extent of the abrupt increase of the number of broad unemployed and broad unemployment rates during the changeover between OHS and LFS. Finally, the use of the QLFS methodology (which required minor revisions) on the LFSs greatly reduced the extent of the abrupt decrease of unemployment aggregates between LFS 2007b and QLFS 2008Q1, thereby improving the comparability of these aggregates across the surveys.

In Chapter 5 poverty and inequality concepts are reviewed, followed by a detailed explanation of the sequential regression multiple imputation (SRMI) technique to deal with households with zero or missing income or expenditure, as well as the derivation of real income, expenditure and consumption variables in each survey. Poverty and inequality trends since the transition are examined in Chapter 6. With regard to poverty, with the exception of AMPS, the poverty trends were very similar across the surveys, that is, poverty increased since the transition, before a downward trend took place since 2000. As far as inequality is concerned, both the levels and trends in the Gini coefficients differed a lot amongst the surveys, as the estimates were very stable in the AMPSs, showed an upward trend in surveys like IESs and GHSs, but first increased until 2000 before a downward trend took place in others (e.g., the two censuses and CS 2007). The levels of inequality also differed when comparing the surveys. The abovementioned poverty and inequality estimates and trends could in part be affected by the various issues discussed in Chapter 3, thus there is a need for careful analysis.

The impact of the number and width of intervals in which income or expenditure data are recorded on poverty and inequality estimates and trends are dealt with in greater detail in Chapter 6 by applying various intervals on the three IESs and NIDS 2008. It is found that the number and width of intervals only had some impact on these estimates and trends in some surveys. The effect of adjusting the survey means in line with the national accounts mean is also investigated. Finally, the application of the cross entropy re-weighting technique did not have any significant impact on the poverty and inequality estimates and trends.

OPSOMMING

Data wat op datum en betroubaar is word vereis om te kan evalueer in watter mate Suid-Afrika sy doelwitte rakende die vermindering van armoede en ongelykheid en die skepping van werkgeleenthede bereik. Sedert die politieke oorgang word verskeie opnamedatastelle gewoonlik vir sulke ontledings gebruik, byvoorbeeld Sensusse, die Gemeenskapsopname van 2007, Inkomste- en Bestedingsopnames, Oktober-huishoudingsopnames, Arbeidsmagopnames, Kwartaallikse Arbeidsmagopnames, Algemene-Huishoudingsopnames, die Nasionale-Inkomste-Dinamika-Studie en die Alle-Media-en-Produkte-opnames.

Weens verskille in steekproef-ontwerp, struktuur van die vraelyste, metodologie om arbeidsmarkstatus te klassifiseer, asook maniere waarop inligting oor inkomste en besteding ingewin is, is hierdie datastelle egter nie ten volle vergelykbaar nie, Gevolglik begin hierdie proefskrif in Hoofstuk 2 om elk van hierdie kwessies in elke opname te ontleed. Inkomste- en bestedingsinligting is in die opnames verskillend ingewin: In die meeste opnames is respondente gevra om aan te dui hoeveel hulle in die verlede bestee of verdien het, maar in die Inkomste- en Bestedingsopname van 2005/2006 is die dagboekmetode gebruik; respondente is in party opnames gevra om die presiese bedrag te vermeld, terwyl hulle in ander opnames die betrokke inkomste- of bestedingsinterval moes aandui; vir eersgenoemde is hulle gevra om òf die enkelbedrag te verklaar, òf hulle moes 'n aantal sub-komponente onderskei; vir intervaldata kan verskillende metodes gebruik word om skattings van die inkomste in elke interval te maak. Dus begin Hoofstuk 3 met 'n oorsig van die voor- en nadele van die verskillende benaderings en 'n bespreking van hoe dit die betroubaarheid en vergelykbaarheid van inkomste- en bestedingsveranderlikes oor die opnames beïnvloed.

In party opnames (bv. die twee sensusse en die Gemeenskapsopname van 2007) dui heelwat huishoudings verkeerdelik aan dat hulle geen inkomste verdien of uitgawes aangaan nie, of hulle spesifiseer nie hoeveel hulle verdien of bestee nie. Ramings van armoede en ongelykheid kan geraak word deur sulke respondent in te sluit of deur hulle uit te laat in die ontledings. Gevolglik word verskeie benaderings in Hoofstuk 3 bespreek om hiermee om te gaan. Omdat opnames vergeleke met die nasionale rekeninge tipies inkomste of besteding onderskat, mag dit die geldigheid van daaruitvoortspruitende armoede- en ongelykheidsramings raak. Gevolglik word argumente vir en teen die aanpsssing van die opname-data om dit in ooreenstemming te bring met die nasionale rekeninge (d.w.s. deur die verdeling na regs te verskuif) bespreek. Ten slotte, omdat die opnamedata streng gesproke

kruissnitdata is en nie ontwerp is vir tydreekse van die arbeidsmag, armoede en ongelykheid nie, word soms aangevoer dat die gewigte van die data herweeg moet word om in ooreenstemming te wees met demografiese en geografiese data soos verkry van die Aktuariële Vereniging van Suid-Afrika en sensusdata. Hierdie kruisentropie herwegingsmetode word in Hoofstuk 3 bespreek. Ten slotte ondersoek die laaste hoofstuk die metodologie vir die bepaling van arbeidsmarkstatus in all die OHS, LFS en QLFS opnames in groter besonderhede, en ook hoe die veranderings oor die verskillende opname-reekse heen dalk die vergelykbaarheid van arbeidsmarkramings deur die jare kan beïnvloed.

Die proefskrif ontleed daarna arbeidsmarktendense sedert die politieke oorgang met gebruik van die Oktober-huishoudingsoponames, Arbeidsmagopnames en Kwartaallikse Arbeidsmagopnames. Beide die arbeidsmag en indiensneming het sedert die transisie toegeneem, maar die toename in indiensneming was onvoldoende om die uitbreiding van die arbeidsmag te absorbeer. Verder het die getal eng-gedefinieerde werkloses tussen 1994 en 2009 verdubbel, en die eng werkloosheidskoers het ‘n toename getoon en in 2003 ‘n toppunt van 30% bereik. Dit het daarna tussen 2004 en 2007 gedaal voordat dit weer in 2008-2009 gestyg het weens die wêreldreseessie. Die toepassing van die kruisentropie-benadering het arbeidsmarktendense nie noemenswaardig beïnvloed nie, wat daarop dui dat hierdie tendense (insluitende die skielike toename in arbeidsmagramings in die oorgang van die Oktober-huishoudingsopname-data na die Arbeidsmarkopname-data) werklik was, of anders plaasgevind het weens veranderings in die opnamevraelyste om respondente se arbeidsmarkstatus beter te probeer bepaal. Verder het die toepassing van die LFS2000b tot LFS 2007B metodologie op die vroeëre opnames die abrupte verlaging in die oorgang tussen die OHS en LFS in die getal breed-gedefineerde werkloses en breë werkloosheidkoerse verminder. Ten slotte het die gebruik van die QLFS-metodologie op die LFS (wat kleiner hersienings benodig het) die abrupte verlaging tussen LFS2007b en QLFS2008Q1 aansienlik verminder, en dus die vergelykbaarheid van hierdie groothede oor die opnames heen verbeter.

In Hoofstuk 5 word eers ‘n oorsig van armoede- en ongelykheidsbegrippe gegee, waarma die sekwensiële-regressie-veelvoudige-imputasie-tegniek in besonderhede bespreek word. Hierdie tegniek word veral gebruik vir gevalle waar huishoudings aandui dat hulle inkomste of besteding nul is, of waar hulle nie antwoord nie. Daar is ook ‘n bespreking van die bepaling van reële inkomste, besteding of verbruiksveranderlikes in elke opname. Armoede- en ongeleykheidstendense word in Hoofstul 6 bespreek. Rakende armoede is daar, met uitsondering van die Alle-Media-en-Produkte-opname, eenstemmigheid dat dit sedert die politieke oorgang eers gestyg het voor dit sedert 2000 begin daal het. Sover dit ongelykheid aanbetref verskil neigings in die Gini-koëffissiënt baie tussen die opnames, want die ramings is stabiel oor die periode vir die Alle-Media-en-Produkte-opname, styg vir die Inkomste- en Bestedingsopname en die Algemene-Huishoudingsopnames, en styg tot 2000 voordat dit afneem in ander opnames (bv. die twee sensusse en die Gemeenskapsopname van 2007). Vlakke van ongelykheid verskil ook tussen die opnames. Deels kan die genoemde tendense in armoede- en ongelykheid dalk toegeskryf word aan die kwessies wat in Hoofstuk 3 bespreek is.

Die effek van die getal en wydte van die intervalle waarin inkomste- en bestedingsdata ingewin word op ramings van armoede en ongelykheid word in meer besonderheid in Hoofstuk 6 bespreek. Deur die toepassing van verskillende intervalle op data van die drie Inkomste- en Bestedingsopnames en die Nasionale-Inkomste-Dinamika-studie word bevind dat die getal en wydte van intervalle 'n beperkte effek op hierdie ramings en tendense het. Verder word gekyk na die effek van die aanpssing van die opname-data om dit in ooreenstemming met die nasionale rekeninge te bring. Ten slotte word getoon dat die gebruik van die kruisentropie-metode nie enige beduidende uitwerking op armoede- en ongeleykheidsramings en -tendense het nie.

ACKNOWLEDGEMENTS

I would not have been able to complete the dissertation without the support of numerous people. First, my supervisor, *Professor Servaas van der Berg* has given me lots of advice, guidance and work opportunities since 2004 (plus the free editing service on my imperfect English on all my research work, including this dissertation!). All my past / present / future achievements would have **never** happened without him.

My parents have always been concerned and worried about me until today. I don't even realise they have aged during the years I worked hard on the dissertation.

I have benefited help from *Professor Murray Leibbrandt*, *Claire Vermaak* and *Dieter von Fintel*, as they helped me understand the multiple imputation concepts and use Stata to impute the data. The assistance from *Professor Martin Wittenberg* and *Nicola Branson* from UCT on the cross entropy re-weighting approach is also appreciated. In addition, I am grateful to receive quick responses from *Mr Piet Alberts* of Statistics South Africa on my questions relating to the labour survey data. I am also fortunate to receive excellent service from the *Stellenbosch University Library*.

I also thank *Professor Lieb Loots* and *Dr Peter Jacobs*, the current and former HODs of UWC respectively, for giving me the precious opportunity to work at UWC since 2006. The lecturing experience helps me realise my lack of knowledge and research done in various areas quicker. Furthermore, the time spent on reading the dissertation and valuable comments made by the *internal and external examiners* are much appreciated.

After running different types of regressions in numerous research projects and on this dissertation, I have learnt an important lesson that it is meaningless to have decent research output on paper, but show an inverse (and significant) relationship between my human capital level as an economist / academic and human morals level as a person behind the good work.

I will continue to work hard after the completion of this dissertation.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AES	Adult equivalence scale
AGR	Actual growth ratio
AMPS	All Media Products Survey
ASGISA	Accelerated and Shared Growth Initiative for South Africa
ASSA	Actuarial Society of South Africa
CBN	Cost of basic needs
CDF	Cumulative density function
CE	Cross entropy
COICOP	Classification of Individual Consumption According to Purpose
CPI	Consumer price index
CS	Community Survey
EAP	Economically active population
EAR	Employment absorption ratio
FEI	Food energy intake
FGT	Foster-Greer-Thorbecke
GDP	Gross domestic product
GEAR	Growth, Employment and Redistribution
GHS	General Household Survey
GIC	Growth incidence curve
ICLS	International Conference of Labour Statisticians
IES	Income and Expenditure Survey
ILO	International Labour Organization
LF	Labour force
LFPR	Labour force participation rate
LFS	Labour Force Survey
MAR	Missing at random
MCAR	Missing completely at random
ME	Maximum entropy
MNAR	Missing not at random
NAAMSA	National Association of Automobile Manufacturers of South Africa
NIDS	National Income Dynamics Study
OHS	October Household Survey
PSLSD	Project for Statistics on Living Standards and Development
QES	Quarterly Employment Statistics
QLFS	Quarterly Labour Force Survey
RDP	Reconstruction and Development Programme

SAARF	South African Advertising Research Foundation
SALDRU	Southern African Labour and Development Research Unit
SAPIA	South African Petroleum Industry Association
SARB	South African Reserve Bank
SARS	South African Revenue Service
SEE	Survey of Employment and Earnings
SRMI	Sequential regression multiple imputation
Stats SA	Statistics South Africa
STC	Standard Trade Classification
TBVC	Transkei-Bophuthatswana-Venda-Ciskei
TGR	Target growth ratio

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 Introduction

South Africa's transition to democracy in 1994 marked a turning point in the history of the country and extended not only political but also social and economic participation to historically disadvantaged groups. This focused policy attention to a much greater extent on issues of poverty and inequality as well as on the labour market and on social delivery mechanisms, as a means of converting political change into social and economic benefits for the wider population. At the same time, and partly in response to the new policy needs, new household level datasets covering the whole country became available that allowed a more thorough analysis of progress in meeting these policy objectives than was possible before. This dissertation focuses on analysing these datasets and extracting information from them in the best possible manner both for valid interpretation of results on labour market, poverty and inequality trends, and to achieve optimal comparability of data across time and between different surveys and censuses. The steps required to arrive at such best estimates and valid comparisons form the core of the study.

As employment growth plays a key role to reduce poverty and inequality, job creation has always been one of the important policy objectives. For instance, the GEAR had the goal of achieving an annual employment growth rate of 6% by creating 400 000 jobs per annum (National Treasury 1996), while ASGISA aimed at reducing the narrow unemployment rate to below 15% by 2014 (National Treasury 2007). The New Growth Path, launched since 2011, aims to create five million jobs over the next 10 years (South African Government 2011). Furthermore, the introduction of legislation such as the Employment Equity Act of 1998 and the Basic Conditions of Employment Act of 1997 helps improving the employment prospects and working conditions of the previously disadvantaged groups.

In order to evaluate the extent to which the country achieved the objectives of poverty and inequality reduction as well as job creation, up-to-date, reliable and comparable data are required. Table 1.1 briefly summarizes the commonly used survey data for labour market, poverty and income distribution analyses (Chapter 2 will discuss each survey in greater detail). Before the transition, the census conducted by Statistics South Africa (Stats SA) was seemingly the only data source available to analyse poverty, inequality and labour market trends. Although the Income and Expenditure Survey (IES) was also a usable dataset, the sample only covered a limited sub-set of households in metropolitan areas of the country. In addition, the 1993 October Household Survey (OHS) excluded the people residing in the TBVC (Transkei-Bophuthatswana-Venda-Ciskei) states from the sample.

Table 1.1: Commonly used surveys for labour market, poverty and inequality analyses in South Africa

Institution	Survey conducted	Year	Frequency
Statistics South Africa (Stats SA)	Census	1996 & 2001	Every five years
	Community Survey (CS)	2007	N/A
	Income and Expenditure Survey (IES)	1995, 2001, 2005/2006	Every five years
	October Household Survey (OHS)	1993 – 1999	Annually
	Labour Force Survey (LFS)	2000 – 2007	Semi-annually
	Quarterly Labour Force Survey (QLFS)	2008 –	Quarterly
	General Household Survey (GHS)	2002 –	Annually
Southern Africa Labour and Development Research Unit (SALDRU)	Project for Statistics on Living Standards and Development (PSLSD)	1993	N/A
	National Income Dynamics Study (NIDS)	2008 –	Every two years
South African Advertising Research Foundation (SAARF)	All Media Products Survey (AMPS)	1993 –	Semi-annually or annually

Since the transition in 1994, a major advance by Stats SA was the improvement of the IES and OHS, as the sample was extended to all areas. In addition, new surveys were conducted, such as the General Household Survey (GHS) introduced in 2002, the Labour Force Survey (LFS) which replaced the OHS since 2000, and the Quarterly Labour Force Survey (QLFS) which replaced the LFS since 2008. The sampling design and questionnaire structure of the aforementioned surveys have also been improved throughout the years.

Institutions other than Stats SA conduct surveys which in turn provide alternative datasets for poverty, inequality and labour market analyses, such as the Project for Statistics on Living Standards and Development (PSLSD) as well as the National Income Dynamics Study (NIDS) conducted by Southern Africa Labour and Development Research Unit (SALDRU). Moreover, although South African Advertising Research Foundation (SAARF) has been conducting the All Media Products Survey (AMPS) since 1975, the data have only been used as an alternative data source for poverty and inequality analyses in recent years.

As the main aim of the OHS, LFS and QLFS is to capture labour market activity in the country, they are the primary data source to be used to derive labour market trends. In South Africa, most studies (e.g., Casale and Posel 2002; Bhorat 2004; Burger and Woolard 2005;

Oosthuizen 2006; Van der Westhuizen, Goga and Oosthuizen 2006) compared the 1995 OHS with the latest available OHS or LFS at the time of writing, before deriving conclusions about the labour market 'trends' for the whole period between surveys. However, this approach is imperfect and could give misleading results, since OHS 1995 is already incomparable with other OHSs/LFSs/QLFSs in many aspects. Hence, a better approach would entail analysing all the available surveys to determine the real trends over time.

However, even after examining all available labour surveys, some abrupt changes in aggregates (e.g., number of labour force participants, the number of employed, number of unemployed) might still be observed during certain years. This could be due to the fact that the OHSs/LFSs/QLFSs are, strictly speaking, cross-sectional national data and are really not designed to be used as a time series, due to changes in the sampling design, shifts in sampling frame, improvement in questionnaire structure and the interview process throughout the years, changes in the methodologies to derive labour market status of the respondents, as well as the adoption of different weighting techniques. This raises concerns about the validity of using these datasets as a time series to examine the labour market trends, and what should be done to improve the reliability and comparability of the abovementioned aggregates across the surveys.

As far as the poverty and inequality trends are concerned, there is an abundance of literature adopting the monetary approach (i.e., per capita income or expenditure variables are used) to derive South Africa's poverty and inequality trends since the advent of democracy. The most commonly used data sets for these analyses are the IESs and censuses (e.g., Simkins 2004; Van der Berg and Louw 2004; Leibbrandt, Levinsohn and McCrary 2005; Hoogeveen and Özler 2006; Leibbrandt, Poswell, Naidoo and Welch 2006). However, in a few recent studies, alternative data sources mentioned above (i.e., NIDS and AMPS) were used (Van der Berg, Louw and Yu 2008, Argent, Finn, Leibbrandt and Woolard 2009 and Leibbrandt, Woolard, Finn and Argent 2010).

Some of the data sets used to derive these trends are problematic in a particular year or in more than one year, which in turn makes it difficult to compare poverty and inequality results across the years. Examples of these problems are high proportion of households with zero or unspecified income in the censuses, and the large decrease in both income and expenditure between the 1995 and 2000 IESs. In addition, different poverty lines were used in the poverty analyses, with the most commonly used poverty line values being R250 per month in 1996

Rand, US$1 a day, US$2 a day, as well as the three poverty lines proposed by Woolard and Leibbrandt (2006), i.e., R211 per month, R322 per month and R593 per month in 2000 prices.

Furthermore, income and expenditure amounts were captured differently amongst the surveys, which in turn could affect the comparability of poverty and inequality trends across the surveys. For example, the recall method was applied in IES 1995 and IES 2000, while the diary method (complemented by the recall method) was introduced in IES 2005/2006; the amounts were captured in either bands or exact terms; the respondents were only asked to declare the overall household income or expenditure amounts in some surveys, while in other surveys the respondents were first asked to declare the income or expenditure from each source, before the overall amount was derived; if the latter method was used to derive the overall amount, the imputed rent variable was included as an income or expenditure item only in some surveys (e.g., IES 2005/2006 and NIDS).

1.2 Research questions

The research questions of the dissertation are as follows:

- How was the information on labour market status as well as income and expenditure captured in each survey?
- What are the issues that could affect the reliability and comparability of the abovementioned information and the subsequent estimates on the labour market, poverty and inequality trends across surveys?
- What has happened to the labour market since the transition until 2009? Using all the available OHS, LFS and QLFS data, the focus is on examining the trends on the labour force (LF) size, labour force participation rate (LFPR), employment, working conditions of the employed, number of unemployed and unemployment rate.
- What are the possible reasons accounting for the changes in the levels and trends of these labour market variables, and what are the likely solutions to improve their reliability and comparability of these aggregates over time?
- What has happened to poverty and inequality levels and trends since the transition and are these trends consistent across different surveys? All the available household data (i.e., Census, IES, PSLSD, NIDS and AMPS data, in addition to the OHS/LFS/QLFS data mentioned above) is used.
- How to address the comparability problems, if any, of the datasets, before the poverty and inequality levels and trends are re-examined?

1.3 Structure of the study

Chapter 2 critically evaluates the sampling methodology, sample size and the questionnaire structure of the surveys under study, namely censuses, Community Survey (CS) 2007, IESs, OHSs/LFSs/QLFSs, GHSs, PSLSD, NIDS and AMPSs. This is followed by an investigation on how the questions relating to income, expenditure and labour market information were asked in each survey. The methodology to derive the labour market status of the respondents in OHSs, LFSs, and QLFSs is also looked at, since these are the surveys with the main aim of capturing the labour market status of people. Finally, the chapter examines whether the sample in some surveys contained a high percentage of households with zero or unspecified household income, expenditure or consumption.

Chapter 3 critically addresses the problems which could affect the comparability of the datasets. The chapter begins by providing a literature review of the pros and cons of using income or expenditure variables for the poverty and inequality analyses, before discussing the advantages and disadvantages of the diary and recall method to collect the income and expenditure information. The income and/or expenditure could be captured either in exact amounts (e.g., IESs) or in bands (e.g., censuses). The merits and drawbacks of each method to capture income and expenditure data are looked at.

If the data are captured in exact amounts, the respondents could be asked to declare the 'one-shot' overall income or expenditure amount (i.e., the single estimate), or asked to declare the amounts from different income and expenditure categories, before the overall amount is derived. With regard to the surveys that capture the income or expenditure information in bands, an important question is the appropriate methodology to use to approximate the income or expenditure amount of the households in each band, as this amount must be derived to make the data continuous, before the per capita variable could be generated for the ensuing poverty and inequality analyses. In addition, the number of bands and the width of the bands, as well households reporting zero or unspecified income or expenditure differ in each survey. Furthermore, it is argued that survey data should be validated or even adjusted against external sources, such as the national accounts income and tax revenue income of the National Treasury. How all these issues could influence the comparability and reliability of poverty and inequality estimates and trends across the surveys, as well as the possible ways to address them are dealt with in Chapter 3.

As mentioned before, since the surveys under study were not designed for time-series comparison, the legitimacy of using these survey datasets as a time series to examine trends in the labour market, poverty and inequality is uncertain (e.g., Branson 2009). Thus, Chapter 3 discusses the cross entropy estimation approach to re-weight the datasets to be consistent with demographic and geographic numbers presented by the Actuarial Society of South Africa (ASSA) model and Census data. Finally, the labour market status derivation methodology in each OHS, LFS and QLFS is examined in greater detail, to investigate how the changes in the methodology could affect the labour market trends.

Chapter 4 begins by providing a literature review of recent studies on the labour market trends, followed by a study on the trends in labour force size and labour force participation rates using all OHS/LFS/QLFS data between 1994 and 2009. Whether the increased 'feminisation of the labour force' as suggested by various studies (Casale and Posel 2002; Casale 2004; Burger and Woolard 2005; Oosthuizen 2006; Van der Westhuizen et al. 2006) took place or not during the period in question is examined. Chapter 4 also discusses employment trends, with specific reference to occupation, industry, skills level and formal/informal sector status of the employed. The issue of jobless growth is also investigated. This is followed by a discussion on the profile of the unemployed as well as the trends in unemployment size and unemployment rates.

However, even examining all available OHSs/LFSs/QLFSs (instead of only comparing one OHS with the latest available LFS to derive the labour market 'trends', as done by many recent studies) does not get rid of the problem that abrupt changes in the labour market aggregates are still observed across some surveys. These abrupt changes could be attributed to what happened to the economy at that time, differences in the labour market status derivation methodology, and the differences in weighting techniques. The last two issues are addressed by using the same labour market status derivation methodology across all surveys, if possible, and by re-weighting the datasets by the cross entropy approach, so as to investigate if the use of a consistent labour market status derivation methodology and weighting techniques have any significant effect in creating a more reliable and valid trend in the labour market variables under study.

Chapter 5 begins by explaining the poverty and inequality concepts and measurements, before providing a detailed literature review of the recent studies on South African poverty and inequality trends. Next, total income and expenditure is derived in each survey. Whether some surveys seriously under-estimated income and/or expenditure, and whether such under-estimation has to do with the issues discussed in Chapters 3 and 4 are also looked at.

In Chapter 6, the three proposed poverty lines by Leibbrandt and Woolard (2006) are used to explore the poverty trends, focusing on the estimates of the Foster-Greer-Thorbecke (FGT) poverty headcount, poverty gap and squared poverty gap ratios. Tests for dominance by means of cumulative density functions (CDFs) are also conducted to investigate if the poverty estimates are sensitive to the poverty lines chosen. Inequality trends are also investigated, focusing on the trends in Gini coefficients as well as racial decomposition of inequality (by means of Theil-L and Theil-T indices to investigate within-race and between-race inequality) across the surveys under the period under study. The poverty and inequality trends derived are then compared with the results from the recent studies. The differences, if any, in the poverty and inequality estimates amongst the surveys are also looked at, before examining how the factors discussed in Chapter 3 could have an impact on the comparability and reliability of these estimates.

Chapter 6 then tries to address the issues discussed in Chapter 3 before re-visiting poverty and inequality estimates and trends. First, as NIDS is the only survey that asked the respondents to report income and expenditure under both the single estimate and aggregation approaches, but only the income and expenditure variables derived from the latter method were used by SALDRU to estimate poverty and income distribution, the chapter investigates if these estimates would be significantly different, if the income and expenditure variables derived from the single estimate method are used instead.

As mentioned before, the income / expenditure / consumption variables in the IES were reported as continuous variables (i.e., respondents declaring the exact amounts earned or spent), so one might ask if the poverty and inequality estimates and trends would differ, had the respondents been asked to report the relevant intervals instead. Hence, the AMPS 2000 intervals are applied on the continuous income variable of IES 2000, but various methods are applied to make the dataset continuous again before poverty and inequality estimates are compared. Furthermore, the Census 1996 and Census 2001 (with relatively wider intervals in the higher-income categories) as well as GHS 2009 (with very few intervals) intervals are also applied on the income variable on IES 2000, so as to investigate if the poverty and inequality estimates are significantly influenced by the number and width of the intervals. The abovementioned intervals are then applied in IES 1995 and IES 2005/2006 to find out if the poverty and inequality estimates and trends notably differ with the application of the different intervals. Finally, the AMPS, GHS and CS 2007 intervals are applied on the NIDS continuous income and expenditure variables to investigate if the aforementioned estimates change significantly.

Chapter 6 then adopts the cross entropy approach to re-weight all the datasets in consideration to investigate if the entropy weights have any significant effect in these trends. Finally, as household surveys might have under-estimated income or expenditure when compared with national accounts income, the survey income or expenditure distribution are adjusted in line with the national accounts mean income before the poverty and inequality trends are re-visited, so as to investigate if the results would differ.

Chapter 7 concludes the dissertation by highlighting trends in labour market, poverty and inequality trends if using the data unadjusted, as well as whether the estimates and trends would differ significantly across the surveys, after the data have been adjusted to deal with the comparability issues. Possible ways to improve the comparability of the datasets further are also suggested.

CHAPTER TWO: THE SOUTH AFRICAN SURVEY DATA

2.1 Introduction

Numerous survey datasets are available to provide income and expenditure information for poverty and inequality analyses, with almost all of the surveys being conducted by Stats SA. However, some data sources provide both income and expenditure data, while others only provide one or the other. In addition, in most of these data sources, the respondents were asked to declare their income or expenditure levels in broad intervals instead of the actual amounts. Furthermore, a large number of households reported zero or unspecified household income or expenditure in some surveys.

As far as the labour market information is concerned, in some surveys (e.g., GHS, IES) only few questions were asked to capture the labour market status of the respondents, as the primary objective of these surveys was not to capture labour market information. Besides, as mentioned in Chapter 1, OHS, LFS and QLFS are the surveys that mainly aim at capturing labour market information, and are thus the primary data sources used to derive labour market trends for the South African economy.

In this chapter, the sampling methodology, as well as how the questions relating to income, expenditure and labour market information were asked in each survey are analysed and critically evaluated, in order to highlight the incomparability issues across the surveys, and lay the groundwork for discussion of corrective steps in the forthcoming chapters.

2.2 Population censuses and Community Survey

Since the political transition, three population censuses were conducted by Stats SA in 1996, 2001 and 2011. As the cabinet decided not to conduct a census in 2006, an information gap between Census 2001 and Census 2011 was created. Hence, a decision was made to conduct the 2007 Community Survey (CS 2007)[1].

2.2.1 Sampling design and sample size

Census 1996 took place in October 1996, and a 10% unit level sample of all households

[1] Strictly speaking, Census 1996 and Census 2001 are not surveys. However, for the remainder of the dissertation, they are referred to as surveys. Also, Census 2011 data are not released yet at the time of writing.

(including special institutions[2], such as hotels, student hostels, churches, prisons, etc.) and all persons as enumerated in the census in South Africa (Stats SA 1998b) was made available.

On the other hand, Census 2001 took place in October 2001. Similarly, a 10% unit level sample was made available, and the sample was drawn as follows (Stats SA 2003c):

- Households: A 10% sample of households in housing units, as well a 10% sample of collective living quarters (both institutional and non-institutional) and the homeless.
- Persons: A sample consisting of all persons in the households and living quarters as well as the homeless, drawn from the sample described above.

In the 10% samples of both censuses, the household records were explicitly stratified according to province and District Council (DC). Within each DC, the records were further implicitly stratified by local authority and enumeration area (EA) type.

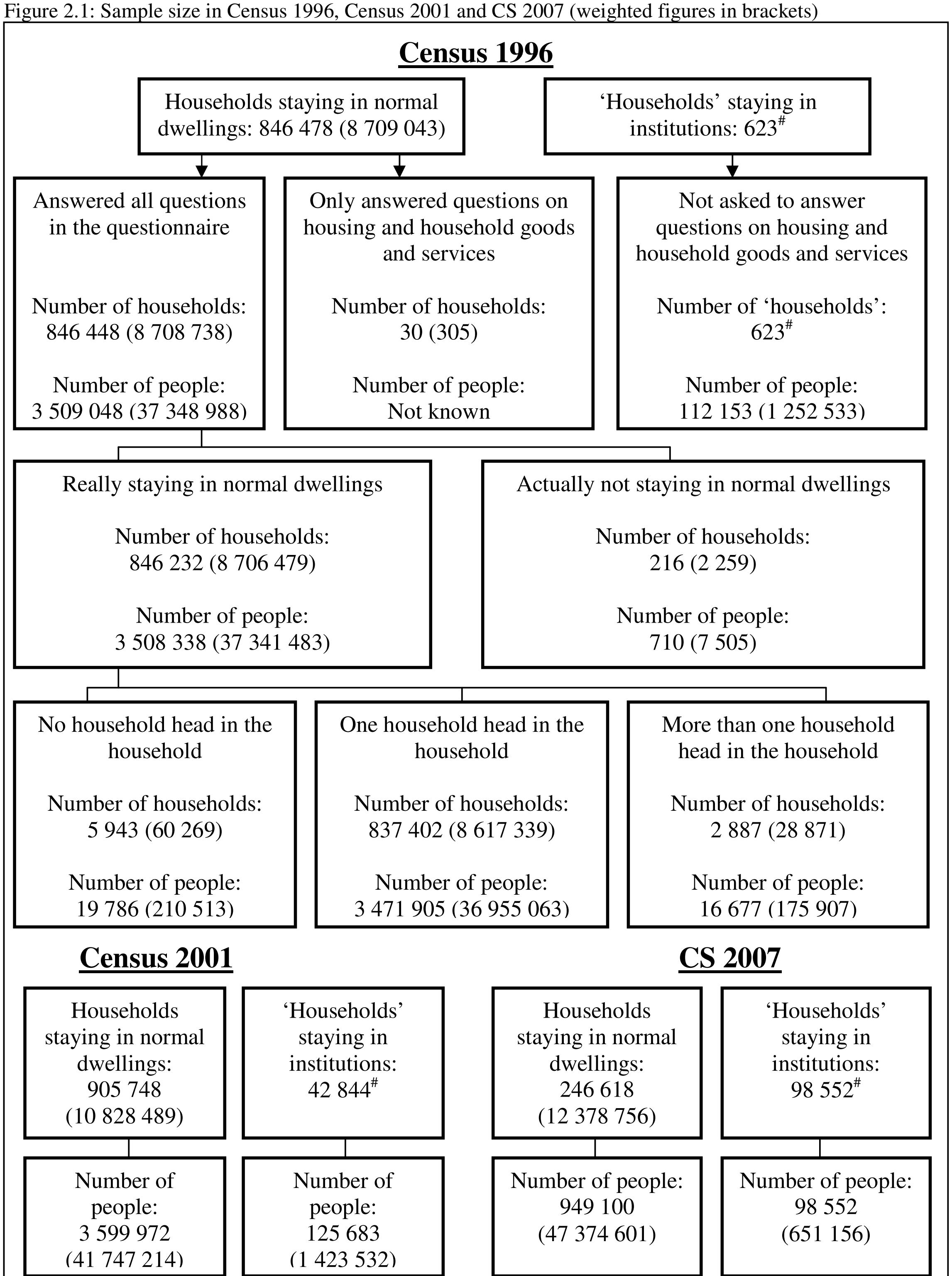
With regard to CS 2007, which took place in February 2007, a two-stage stratified random sampling process was adopted (Stats SA 2008f). In the first stage, each municipality was regarded as an explicit stratum and a systematic sampling method was adopted to select EAs within each municipality, with the EAs being ordered by geographic and EA types. The second stage involved the selection of dwelling units. Such selection was based on a fixed proportion of 10% of the total listed dwellings in an EA. All households within the selected dwelling units were covered. Besides, there was no replacement of vacant dwellings, refusals or non-contacts because of their impact on the probability of selection. Hence, Stats SA made concerted efforts to improve the response rates by means of multiple visits.

In the 10% sample of Census 1996, 846 478 households stayed in normal dwellings, according to their answers in the dwelling type question in the household-level section (i.e., Section B of the questionnaire)[3]. However, 30 of them only answered questions in Section B but did not take part in the person-level section (i.e., Section A)[4]. In other words, only the people from the remaining 846 448 households staying in normal dwellings took part in all sections of the questionnaire. However, of these 846 448 households, 216 contradicted their dwelling type answers in Section B by claiming they actually stayed at institutions in Section A. Therefore, the correct number of households staying in normal dwellings was 846 232.

[2] For the remainder of the dissertation, households that did not reside at institutions at the time of the survey will be referred to as “households living in normal dwellings”.
[3] The questionnaire structure will be discussed in greater detail in Section 2.2.2.
[4] These 30 households will be excluded from the analyses for the remainder of the dissertation.

In Census 2001 10% sample and CS 2007, 905 748 and 246 618 households staying in normal dwellings took part in the survey respectively.

Figure 2.1: Sample size in Census 1996, Census 2001 and CS 2007 (weighted figures in brackets)



It is not possible to derive the weighted number of "households" staying in institutions in all three surveys, because there was no household weight value given to these "households". Only person weight was available for people staying in institutions. Hence, only the unweighted number of "households" staying in institutions (623 in Census 1996 and 98 552 in CS 2007) could be derived.

Figure 2.1 above summarizes sample size in each survey, with the weighted figures in brackets. For the remainder of the study, people from households staying in normal dwellings are the focus of the analyses.

2.2.2 The questionnaire

2.2.2.1 Questionnaire structure

In Census 1996 and Census 2001, there were two sections in the questionnaire. Section A asked questions on demographics, migration status, educational attainment, labour market status, economic activities and personal income. In Section B, information on housing (Example: dwelling type, ownership of dwelling, number of rooms, sharing of the same room by more than one person, etc.) and household goods and services (Example: water access, landline telephone in dwelling, sanitation, energy/fuel, refuse removal, ownership of goods like television and computer) was captured. Similar questions were asked in CS 2007, with the addition that questions on receipt of social grants were asked.

2.2.2.2 Labour market status questions

With regard to the labour market information, in Census 1996, questions 17 and 18 (which contained three sub-questions) of Section A were asked to identify the broad[5] labour market status of the respondent. In addition, question 19, which was split into seven sub-questions, was asked to capture the working conditions of the employed, such as their occupation, industry, place of work, whether they were self-employed or employees, as well as whether they worked full-time or not.

In Census 2001, question 18 of Section A, which was split into four sub-questions, was asked to capture both the narrow and broad labour market status of the respondent. In addition, question 19, which contained seven sub-questions, was asked to capture the same information on the working conditions of the employed as in Census 1996, with the addition that the information on the work hours of the employed in the past seven days was also captured in Census 2001.

Section E (i.e., questions 30 – 38) of CS 2007 helped capture the narrow labour market status of the respondents. In addition to asking the similar questions as in Census 2001, CS 2007 also asked questions on the respondents' action (if any) to seek work if they were not employed at the time of the survey, and the formal/informal sector status of the employed.

[5] The difference between the narrow and broad labour market status will be discussed in Section 2.4.2.2.

2.2.2.3 Household income and expenditure questions

In all three surveys, household expenditure was not captured, while each member of the household was asked to declare his/her relevant personal income category. Furthermore, in 1996, the household was asked to declare the additional income and remittances received[6].

In Census 1996, the three income questions were asked as follows:

- Personal income (Question 20, Section A): "Think of the past year (1 October 1995 to 30 September 1996) and the money each person received. Please indicate this person's income category before tax. Answer this question by indicating each person's weekly, monthly or annual income. Include all sources of income, for example housing loan subsidies, bonuses, allowances such as car allowances and investment income. If this person receives a pension or disability grant, please include this amount."
- Additional income (Question 1.1, Section B): "Think of any additional money that this household generates, and that has not been included in the previous section (For example, the sale of home-grown produce or home-brewed beer or cattle or the rental of property. Please indicate this total amount, if anything, during the past year. (1 October 1995 – 30 September 1996). If none enter '0'."
- Remittances received (Question 1.2, Section B): "If this household receives any remittances or payments (for example money sent back home by someone working or living elsewhere or alimony), please indicate the total received during the past year. (1 October 1995 – 30 September 1996). If none enter '0'."

All three income variables were recorded in exactly the same intervals, as shown in the first column of Table 2.1. Next, Stats SA derived the personal income amount for each person (See the second column of Table 2.1), before the personal income amounts of all members of the household were added together[7]. The additional household income and household remittances

[6] For the remainder of the study, these two income variables will be referred to as 'additional household income' and 'household remittances received' respectively.

[7] These amounts were derived as follows (Stats SA 1998b):

- Persons claiming they had zero monthly income were not adjusted
- For the first category among those with incomes (R1 – R200), the amount was approximated two-thirds of the top cut-off point of this bracket, i.e., R200 × 2/3 = R133.33.
- For the second category (R201 – R500), the amount was the midpoint of the class interval, i.e., (R201 + R500)/2 = R350
- For the last category (R30 001 or more), the amount was twice the cut-off point of the second last class (R16 001 – R30 000), i.e., R30 000 × 2 = R60 000.
- For the other classes, the amount was calculated as the logarithmic mean of the top and bottom of the given interval, e.g., looking at the R501 – R1 000 category, the amount was equal to: exp^([ln(R501) + ln(R1 000)]/2) = R707.17.

amounts were estimated in exactly the same way, and were added to the total personal income amounts of all members of the household, before the household income amount was derived. Finally, the result for each household was reallocated into the relevant household income category. The household income categories were exactly the same as the personal income categories (i.e., the first column of Table 2.1).

Table 2.1: Derived monthly personal income, additional household income, and household remittances amounts, Census 1996

	Derived personal income / additional household income / household remittance amount
1: None	R0.00
2: R1 – R200	R133.33
3: R201 – R500	R350.00
4: R501 – R1 000	R707.17
5: R1 001 – R1 500	R1 224.83
6: R1 501 – R2 500	R1 936.58
7: R2 501 – R3 500	R2 958.08
8: R3 501 – R4 500	R3 968.67
9: R4 501 – R6 000	R5 196.17
10: R6 001 – R8 000	R6 928.25
11: R8 001 – R11 000	R9 380.92
12: R11 001 – R16 000	R13 266.58
13: R16 001 – R30 000	R21 909.00
14: R30 001 or more	R60 000.00
99: Unspecified	N/A

As some respondents did not specify their personal income, and some households did not specify the additional household income and household remittances, Stats SA adopted the following three rules when the household income was derived (Stats SA 1998b):

- If personal income was unspecified for a member of the household aged under 15 years, then the personal income for this child was set to R0.
- If a member of the household aged 15 years or older had unspecified personal income, it remained unspecified, and the household income was taken to be unspecified as well, because there was not enough information for the estimate to be reliable.
- If additional household income or remittances received were unspecified, they were set to R0.

Table 2.2 gives three examples of the derivation of household income amount and category.

Table 2.2: Derivation of monthly household income amount and category, Census 1996

Household A	
Personal income of member #1 (Aged 20 years):	4: R501 – R1 000
Personal income of member #2 (Aged 12 years):	99: Unspecified
Personal income of member #3 (Aged 40 years):	13: R16 001 – R30 000
Additional household income:	4: R501 – R1 000
Receipt of remittances:	4: R501 – R1 000
Household income amount:	R24 030.51 (707.17 + 0 + 21 909.00 + 707.17 + 707.17)
Household income category:	13: R16 001 – R30 000
Household B	
Personal income of member #1 (Aged 20 years):	4: R501 – R1 000
Personal income of member #2 (Aged 16 years):	99: Unspecified
Personal income of member #3 (Aged 40 years):	13: R16 001 – R30 000
Additional household income:	4: R501 – R1 000
Receipt of remittances:	4: R501 – R1 000
Household income amount:	Unspecified
Household income category:	99: Unspecified
Household C	
Personal income of member #1 (Aged 20 years):	4: R501 – R1 000
Personal income of member #2 (Aged 12 years):	99: Unspecified
Personal income of member #3 (Aged 40 years):	13: R16 001 – R30 000
Additional household income:	99: Unspecified
Receipt of remittances:	99: Unspecified
Household income amount:	R22 616.17 (707.17 + 0 + 21 909.00 + 0 + 0)
Household income category:	13: R16 001 – R30 000

However, when analysing the household income variable derived by Stats SA in greater detail, it was found that the three rules mentioned above were not applied in some households:

- 295 541 households did not contain any member aged 15 years or above with unspecified personal income. Therefore, these households should have had specified household income. However, it can be seen from Table 2.3 that these households strangely had unspecified household income.
- 724 894 households (The sum of all the values in the third column of Table 2.3, excluding the value 276 423) had at least 1 member aged 15 years or above with unspecified personal income. Thus, according to the rules mentioned above, these households should have unspecified household income. However, the results from Table 2.3 show that these households still had specified income.

Hence, it was decided to apply the three rules to derive the household income variable again, and the results, which are different from using the Stats SA household income variable, are presented in Table 2.4. Thus, it seems that the household income variable derived originally by Stats SA is not accurate. For the remainder of the dissertation, the 1996 household income variable derived by the author will be used, unless stated otherwise.

Table 2.3: Number of households in each monthly household income category, Census 1996

Household income (Derived by Stats SA)	With zero members aged 15 years or above with unspecified personal income	With at least one member aged 15 years or above with unspecified personal income	Total
1: None	1 070 378	149 396	1 219 774
2: R1 – R200	636 703	51 675	688 378
3: R201 – R500	1 260 775	91 757	1 352 532
4: R501 – R1 000	1 095 109	98 430	1 193 539
5: R1 001 – R1 500	773 272	90 474	863 746
6: R1 501 – R2 500	749 347	80 215	829 562
7: R2 501 – R3 500	427 364	45 724	473 088
8: R3 501 – R4 500	313 351	30 920	344 271
9: R4 501 – R6 000	334 233	31 090	365 323
10: R6 001 – R8 000	230 351	19 567	249 918
11: R8 001 – R11 000	237 913	16 386	254 299
12: R11 001 – R16 000	147 409	10 111	157 520
13: R16 001 – R30 000	102 815	6 785	109 600
14: R30 001 or more	30 601	2 364	32 965
99: Unspecified	295 541	276 423	571 964
	7 705 162	1 001 317	8 706 479

Table 2.4: Number of households in each monthly household income category as derived by Stats SA and the author respectively, Census 1996

	Household income (Derived by Stats SA)		Household income (Derived by the author)	
1: None	1 219 774	14.0%	1 129 419	13.0%
2: R1 – R200	688 378	7.9%	558 158	6.4%
3: R201 – R500	1 352 532	15.5%	1 402 548	16.1%
4: R501 – R1 000	1 193 539	13.7%	1 074 861	12.3%
5: R1 001 – R1 500	863 746	9.9%	848 328	9.7%
6: R1 501 – R2 500	829 562	9.5%	777 787	8.9%
7: R2 501 – R3 500	473 088	5.4%	435 253	5.0%
8: R3 501 – R4 500	344 271	4.0%	333 118	3.8%
9: R4 501 – R6 000	365 323	4.2%	354 256	4.1%
10: R6 001 – R8 000	249 918	2.9%	241 198	2.8%
11: R8 001 – R11 000	254 299	2.9%	249 667	2.9%
12: R11 001 – R16 000	157 520	1.8%	157 704	1.8%
13: R16 001 – R30 000	109 600	1.3%	109 974	1.3%
14: R30 001 or more	32 965	0.4%	32 891	0.4%
99: Unspecified	571 964	6.6%	1 001 317	11.5%
	8 706 479	100.0%	8 706 479	100.0%

In Census 2001, household income was derived by adding the derived personal income amounts of all household members, with the personal income question asked as follows (Question 22, Section A): "What is the income category that best describes the gross income of this person before tax?". This income variable was recorded in different intervals than in Census 1996 (See the first column of Table 2.5). The personal income amount, as shown in the second column of Table 2.5, was estimated using the same method as in Census 1996[8].

[8] These amounts were derived as follows (Stats SA 2003c):

Table 2.5: Derived monthly personal income amounts, Census 2001 and CS 2007

	Derived personal income amount
1: None	R0
2: R1 – R400	R266.67
3: R401 – R800	R600.00
4: R801 – R1 600	R1 131.33
5: R1 601 – R3 200	R2 262.75
6: R3 201 – R6 400	R4 525.50
7: R6 401 – R12 800	R9 051.00
8: R12 801 – R25 600	R18 101.92
9: R25 601 – R51 200	R36 203.83
10: R51 201 – R102 400	R72 407.75
11: R102 401 – R204 800	R61 482.17
12: R204 801 or more	R409 600.00
13: Unspecified	N/A

15.6% of respondents had unspecified personal income in Census 2001 (See the third column of Table 2.6). However, Stats SA applied the hot deck imputation method[9] to impute their personal income category, and the results are shown in the last three columns of Table 2.6.

Table 2.6: Number of people in each monthly personal income category before and after hot deck imputation, Census 2001

	Before hot deck imputation		After hot deck imputation	
1: None	23 434 110	56.1%	28 712 005	68.8%
2: R1 – R400	2 046 913	4.9%	2 310 421	5.5%
3: R401 – R800	3 663 976	8.8%	4 028 173	9.6%
4: R801 – R1 600	2 008 797	4.8%	2 183 074	5.2%
5: R1 601 – R3 200	1 706 388	4.1%	1 876 788	4.5%
6: R3 201 – R6 400	1 263 542	3.0%	1 404 969	3.4%
7: R6 401 – R12 800	677 332	1.6%	759 272	1.8%
8: R12 801 – R25 600	256 999	0.6%	289 125	0.7%
9: R25 601 – R51 200	89 543	0.2%	99 929	0.2%
10: R51 201 – R102 400	35 182	0.1%	40 058	0.1%
11: R102 401 – R204 800	25 877	0.1%	32 101	0.1%
12: R204 801 or more	9 859	0.0%	11 299	0.0%
13: Unspecified	6 528 696	15.6%	0	0.0%
	41 747 214	100.0%	41 747 214	100.0%

After hot deck imputation, everyone had a specified personal income. Next, household income was derived by summing the personal income amounts of all members in the

- Persons claiming they had zero monthly income were not adjusted
- For the first category among those with incomes (R1 – R400), the amount was approximated two-thirds of the top cut-off point of this bracket, i.e., R400 × 2/3 = R266.67.
- For the second category (R401 – R800), the amount was the midpoint of the class interval, i.e., (R401 + R800)/2 = R600
- For the last category (R204 801 or more), the amount was twice the cut-off point of the second last class (R102 401 – R204 800), i.e., R204 800 × 2 = R409 600.
- For the other classes, the amount was calculated as the logarithmic mean of the top and bottom of the given interval, e.g., looking at the R801 – R1 600 category, the amount was equal to: exp^([ln(R801) + ln(R1 600)]/2) = R1 131.33.

[9] The hot deck imputation method as well as some other imputation methods will be discussed in Chapter 3.

household. For example, if a household contained three members, and the monthly personal income categories of the three persons (after hot deck imputations) were “1: R0”, “2: R1 – R400” and “10: R51 201 – R102 400” respectively, then the household income amounted to R72 674.42 (R0 + R266.67 + R72 407.75). Finally, the result for each household was reallocated into the relevant household income category. The household income categories were exactly the same as the personal income categories. Table 2.7 shows that 16.4% of households had unspecified household income before hot deck imputation was applied on unspecified personal income.

Table 2.7: Number of households in each monthly household income category, Census 2001

	Before hot deck imputation		**After hot deck imputation**	
1: None	2 274 882	21.0%	2 546 711	23.5%
2: R1 – R400	774 583	7.2%	877 609	8.1%
3: R401 – R800	1 686 640	15.6%	1 927 235	17.8%
4: R801 – R1 600	1 437 798	13.3%	1 728 296	16.0%
5: R1 601 – R3 200	1 119 402	10.3%	1 403 207	13.0%
6: R3 201 – R6 400	759 920	7.0%	989 325	9.1%
7: R6 401 – R12 800	529 351	4.9%	710 802	6.6%
8: R12 801 – R25 600	302 734	2.8%	412 495	3.8%
9: R25 601 – R51 200	107 869	1.0%	146 940	1.4%
10: R51 201 – R102 400	29 814	0.3%	41 814	0.4%
11: R102 401 – R204 800	19 051	0.2%	28 256	0.3%
12: R204 801 or more	11 038	0.1%	15 799	0.1%
13: Unspecified	1 775 407	16.4%	0	0.0%
	10 828 489	100.0%	10 828 489	100.0%

In CS 2007, Stats SA derived the household income by summing the personal income amounts of all members in the household, with the personal income question being asked as (Question 52, Section G): “What is the income category that best describes the gross monthly or annual income of (the person) before deductions and including all sources of income?” With regard to the derivation of the personal income amount, it was estimated using exactly the same method as in Census 2001 (See footnote 8), and since the income categories between Census 2001 and CS 2007 were exactly the same in nominal Rand terms, the derived personal income amounts in each category were also the same across the two surveys (Table 2.5). As far as the derivation of household income is concerned, in cases where there was unspecified personal income for any member, regardless of age, the household income was set to be unspecified as well (Stats SA, 2008f).

Finally, similar to the two censuses, the result for each household was reallocated into the relevant household income category. The household income categories were exactly the same as the personal income categories. Table 2.8 presents the results.

Table 2.8: Number of households in each monthly household income category, CS 2007

1: None	1 011 941	8.2%
2: R1 – R400	617 704	5.0%
3: R401 – R800	1 108 092	9.0%
4: R801 – R1 600	2 343 212	18.9%
5: R1 601 – R3 200	2 361 470	19.1%
6: R3 201 – R6 400	1 416 124	11.4%
7: R6 401 – R12 800	943 714	7.6%
8: R12 801 – R25 600	659 274	5.3%
9: R25 601 – R51 200	352 141	2.8%
10: R51 201 – R102 400	116 839	0.9%
11: R102 401 – R204 800	40 259	0.3%
12: R204 801 or more	28 790	0.2%
13: Unspecified	1 379 196	11.1%
	12 378 756	100.0%

To conclude, information on household expenditure was not captured in all three surveys, while the household income variable was derived differently across the three surveys. However, the 1996 household income variable was not derived correctly by Stats SA. Furthermore, all three surveys contain a high proportion of households with zero or unspecified income (before hot deck imputation, in the case of Census 2001), and the number of intervals as well as the width of each interval in real terms in CS 2007 are not comparable to those in Census 1996 and Census 2001. Chapter 3 will examine how these issues would affect the comparability and reliability of poverty and inequality estimates across the surveys, and the possible ways to address them.

2.3 Income and expenditure surveys (IESs)

The IES is also conducted by Stats SA. Since the transition, three IESs took place, in October 1995, October 2000, and between September 2005 and August 2006[10]. The primary objective of the IES is to collect and provide information on income and expenditure patterns of a representative sample of households, so as to update the basket of goods and services required for the compilation of the consumer price index (CPI). Nonetheless, these surveys have also become an important source of information for deriving poverty and inequality estimates and trends.

Table 2.9 presents the general information on the three IESs, such as the sample size, sampling frame, linkages with other surveys, survey period, number of questionnaires,

[10] These three surveys are countrywide surveys, covering metropolitan, urban and rural areas. Thus, they are not comparable with their predecessors, since the latter surveys only covered a more limited sub-set of households in metropolitan areas of the country.

number of visits per household by the fieldworkers, as well as the methodology to capture income and expenditure data. Some of them will be discussed in greater detail in this section.

Table 2.9: General information on the three IESs

	IES 1995	**IES 2000**	**IES 2005/2006**
Census used as a frame for drawing the sample for IES	Census 1991	Census 1996	Census 2001
Linkage with other surveys, if any	OHS 1995	LFS 2000 September	None
Questionnaires	One main questionnaire	One main questionnaire	One main questionnaire + Four weekly diaries
Methodology to capture data	Recall: Income and expenditure on non-durable items, semi-durable items, durable items and services	Recall: Income and expenditure on non-durable items, semi-durable items, durable items and services	Recall: Income and expenditure on semi-durable items, durable items and services Diary: Expenditure on non-durable items, semi-durable items and durable items
Number of visits per household	One	One	Minimum: Six# Maximum: Nine#
Survey period	Oct 1995	Oct 2000	Sep 2005 – Aug 2006
Sample size (Number of households)	29 582	26 263	21 144
Sample size in each group (IES 2005/2006 only)	N/A	N/A	Group 1: 5 253 ○ Sep 2005: 1 796 ○ Oct 2005: 1 729 ○ Nov 2005: 1 728 Group 2: 5 230 ○ Dec 2005: 1 786 ○ Jan 2006: 1 809 ○ Feb 2006: 1 635 Group 3: 5 356 ○ Mar 2006: 1 900 ○ Apr 2006: 1 717 ○ May 2006: 1 739 Group 4: 5 305 ○ Jun 2006: 1 785 ○ Jul 2006: 1 841 ○ Aug 2006: 1 679
Classification of expenditure items	Standard Trade Classification (STC)	Standard Trade Classification (STC)	Classification of Individual Consumption According to Purpose (COICOP)
Others things to take note of	Only a maximum number of 10 members per household could take part in the survey		

If the household only completed one weekly diary, then the number of visits by the fieldworkers would be six (i.e., five visits to ask questions from the main questionnaire and one visit to collect the weekly diary). However, if the household completed all four weekly diaries, then the number of visits by the fieldworkers would be nine.

2.3.1 Sampling design and sample size

In IES 1995, information was obtained from 29 582 households[11], with most of them being linked to OHS 1995[12]. In total, 3 000 enumerator areas (EAs) were drawn for the sample, and 10 households were visited in each EA. The sample was stratified by race, province and area type. In addition, Census 1991 was used as a frame for drawing the sample (Stats SA 1997b). The survey took place in October 1995, and data collection consisted of an extensive interview using a pre-coded main questionnaire. Only a maximum number of 10 members per household were allowed to take part in the survey, and only a maximum number of five members per household (household head, his/her spouse and any other three members) were allowed to answer the questions in the income section.

In IES 2000, Census 1996 was used for drawing the sample. Information was obtained from 26 263 households, and most of these households were linked to the September 2000 LFS[13]. Altogether, 3 000 primary sampling units (PSUs) were drawn for the sample, and explicit stratification of the PSUs was done by province and area type. Within each explicit stratum, the PSUs were implicitly stratified by District Council (DC) and Magisterial District (MD). Next, a systematic sample of about 10 dwelling units was drawn from each PSU (Stats SA 2002c). The survey took place in October 2000, and as in IES 1995, data collection in IES 2000 consisted of an extensive interview using a pre-coded main questionnaire.

In IES 2005/2006, a newly designed sample consisting of approximately 3 000 PSUs, based on Census 2001 enumeration areas, was used as the sampling frame. These 3 000 PSUs were representatively divided into four quarterly allocations of 750 each. A random sample of 250 PSUs was selected every month within each quarterly allocation. Next, eight dwelling units were chosen from each of the sampled PSUs for fieldwork. The aim of this process was to ensure that the sample was evenly spread over the 12-month period, while it remained nationally representative in each quarter (Stats SA 2008d). During the 12 months of data collection, 25 192 households were covered in total, but 4 048 of them were rejected later and excluded from the final data for numerous reasons, which will be discussed in Section 2.3.2.3. Hence, the final sample size in IES 2005/2006 was 21 144 households.

[11] In all three IESs, a household was taken to include all people who lived together for at least 4 days a week at the time of the survey, and babies were included.
[12] Of the 29 582 households interviewed in IES 1995, 28 585 households also took part in OHS 1995.
[13] Of the 26 263 households interviewed in IES 2000, 26 226 households also took part in the 2000 September LFS.

The IES 2005/2006 survey was conducted from September 2005 to August 2006, with sampled households participating for one month and new sub-samples of households starting every month. Data collection consisted of an extensive interview using a pre-coded main questionnaire, which was split and conducted on five separate visits during the survey month. In addition, the households were required to record all its acquisitions in diaries during the survey month. A separate diary was used for each of the four weeks of the survey month, and the diary was collected on a weekly basis.

2.3.2 The questionnaire

2.3.2.1 Questionnaire structure

In IES 1995, there was a short section at the beginning asking questions on the demographic characteristics (e.g., gender, age, race) and the profession (if employed) of each member of the household, the main income source of the household head, as well as particulars of housing. It was followed by a long section that asked questions on the household expenditure on various items, as well as income from different sources.

At the beginning of the IES 2000 questionnaire, a short section was devoted to capture the demographic information of each member of the household. In addition, each member of the household was asked if he/she worked in the last seven days, as well as his/her occupation and industry, if employed. This was followed by another short section that asked questions on the area of purchase of goods and services, as well as information regarding dwellings (e.g., dwelling type, water source, sanitation facility, etc.). After that, as in IES 1995, there was a long section asking questions relating to household income and expenditure.

With regard to IES 2005/2006, as in IES 2000, questions on demographic information of each member of the household, area of purchase of goods and services, as well as information regarding dwellings were asked first, before the household income and expenditure questions were asked. However, looking at the section that captured demographic information, the respondent was no longer asked to declare his/her work status in the past seven days as in IES 2000. Instead, he/she was asked to declare his/her main income source (i.e., Question 1.1a of Section 1), with “salaries and wages” as one of the categories.

2.3.2.2 Labour market status questions

As the name of the survey suggests, the focus of the IES is to capture information on income and expenditure. Hence, each IES only asked a limited number of questions relating to labour market information, as discussed in Section 2.3.2.1. It was not possible to derive the labour market status (in both narrow and broad terms) of the respondents using these questions.

2.3.2.3 Household income and expenditure questions

In both IES 1995 and IES 2000, there was only one main questionnaire that asked questions on both income and expenditure. Table 2.10 shows how the expenditure items were categorized by means of the Standard Trade Classification (STC) approach. In the expenditure section, the households were asked to declare the expenditure during the month prior to the survey for some items and the expenditure for the 12 months prior to the survey for other items. The former was multiplied by 12 before they were converted into annual figures. The only exception was the items under reading matter and stationery (i.e., category 17), as the household was given the option to declare this expenditure on a weekly, monthly or annual basis. The weekly expenditure and monthly expenditure amounts were multiplied by 52 and 12 respectively so as to become annualized figures. Finally, the sum of the expenditure on all items from the first twenty categories (i.e., category 21 – debt – was excluded) was equal to the total household annual expenditure.

Table 2.10: Categorization of expenditure items, IES 1995 and IES 2000

	Expenditure category / sub-category	Weekly/Monthly/Annual
(1)	Housing	
o	Regular housing cost	Monthly
o	Occasional housing cost	Annual
(2)	Domestic workers	
o	Domestic workers	Monthly
(3)	Food	
o	Cereal	Monthly
o	Meat	Monthly
o	Fish and other seafood	Monthly
o	Butter, fats, oils and margarine	Monthly
o	Milk, milk substitutes, cheese and eggs	Monthly
o	Vegetables	Monthly
o	Fruit and nuts	Monthly
o	Sugar, sugar products and sweeteners	Monthly
o	Syrup, jam and related products	Monthly
o	Coffee, tea and cocoa	Monthly
o	Other food products	Monthly
o	Meals/Snacks purchased/consumed away from home	Monthly
o	Baby food	Monthly
(4)	Beverages	
o	Non-alcoholic beverages consumed away from home	Monthly
o	Non-alcoholic beverages consumed at home	Monthly
o	Alcoholic drinks consumed away from home	Monthly
o	Alcoholic drinks consumed at home	Monthly
(5)	Cigarettes, cigars, tobacco, etc. and smokers' requisites	
o	Cigarettes, cigars, tobacco, etc. and smokers' requisites	Monthly
(6)	Personal care	
o	Personal care	Monthly

Table 2.10: Continued

Expenditure category / sub-category		Weekly/Monthly/Annual
(7)	Other household consumer goods	
o	Other household consumer goods	Monthly
(8)	Household services	
o	Household services	Monthly
(9)	Household fuel	
o	Household fuel	Monthly
(10)	Clothing and footwear	
o	Clothing	Annual
o	Footwear	Annual
o	Made-up clothing	Annual
(11)	Furniture/Equipment	
o	Furniture, fixtures and floor coverings	Annual
o	Household textiles	Annual
o	Appliances	Annual
o	Other household equipment	Annual
(12)	Health services	
o	Medical cost of members of medical aid	Annual
o	Medical cost of non-members of medical aid	Annual
(13)	Transport	
o	Cost of private transport purchased	Annual
o	Running costs of private transport	Annual
o	Cost of public and hired transport for work/school purpose	Annual
o	Cost of public and hired transport for holiday	Annual
(14)	Computer and telecommunication equipment	
o	Computer and telecommunication equipment	Annual
(15)	Communication for household purposes	
o	Communication for household purposes	Annual
(16)	Education	
o	Educational expenditure borne by households	Annual
o	Educational expenditure covered by grants	Annual
(17)	Reading matter and stationery	
o	Reading matter and stationery	Weekly or Monthly or Annual
(18)	Recreation, entertainment and sports	
o	Instruments, equipment and accessories	Annual
o	Other goods for recreation/entertainment/sports	Annual
o	Licenses, rental and other service charges	Annual
(19)	Miscellaneous expenditure	
o	Jewellery, handbags, sunglasses	Annual
o	Membership fees, remittances, etc.	Annual
o	Income tax (PAYE/SITE + other payments – refunds received)	Annual
o	Finance and insurance	Annual
o	Other expenditure (gambling, funeral, legal fees, etc.)	Annual
o	Net loss	Annual
(20)	Expenditure on own harvest/livestock	
o	Produce	Annual
o	Livestock	Annual
o	Input cost	Annual

A few expenditure items were asked for the first time in IES 2000, for example, expenditure on pre-cooked frozen meats, prepaid salads, and petrol/diesel for non-transport-related household use. In contrast, expenditure on a few items that were asked in IES 1995 were dropped in IES 2000, for example, expenditure on towels and face cloths, insurance premiums paid on accident policies, and expenditure on dry cleaning services. However, the expenditure on these items was quite small and will only have negligible influence on composition of total household expenditure between the two surveys.

With regard to derivation of household annual income, each person in the household was required to declare his annual regular income and irregular income from different sources, as shown in Table 2.11. However, as mentioned in Section 2.3.1, only a maximum number of five members per household (household head, his/her spouse and any other three members) were asked to answer the questions in the income section in IES 1995. Total household annual income was equal to the sum of the personal regular income and irregular income of the household members.

Table 2.11: Categorization of income items, IES 1995 and IES 2000

	Regular income
(1)	Salaries and wages
○	Salaries and wages from normal hours worked
○	Bonuses and income from overtime
○	Commission and director's fees
○	Part-time work and cash allowances in respect of transport, housing and clothing
(2)	Net profit from business/professional practice/farming on a full-time or part-time basis
(3)	Net income from letting of fixed property
(4)	Royalties
(5)	Interest received on savings/deposits/etc.
(6)	Dividends received on shares other than building society shares
(7)	Receipts from pension/social welfare grants/annuity funds
○	Pension from employment before retirement
○	Annuity and similar recurring receipts resulting from own investment
○	Old-age and war pensions
○	Disability grants
○	Family and other allowances
○	Income from Workmen's Compensation, Unemployment Insurance Fund and similar funds
(8)	Alimony, maintenance and similar allowances from family members living elsewhere
(9)	Regular allowances received from family members living elsewhere
	Irregular income
(1)	Net income from hobbies, side-lines and part-time activities
(2)	Income derived from sale of motor vehicles, fixed property, as well as all other personal property and second-hand goods
(3)	Payments received from borders and other members of the household
(4)	Value of goods and services received by virtue of your occupation and shown as expenditure in the questionnaire
○	Housing (value of subsidies, reduced interest rates and rent, etc.)
○	Transport (value of company transport for private use, reduced air and train fares, etc.)
○	Pension, provident, medical and annuity funds
○	Other
(5)	Gratuities and other lump-sum payments
○	Lump sum resulting from your own employment before retirement
○	Endowment policies and other similar lump sums
○	Lump sums received from the Workmen's Compensation, unemployment insurance, etc.
○	Life insurance and inheritances received

Table 2.11: Continued

Irregular income	
(6)	Claims
o	Funeral funds, including funds' contributions to funeral expenses
o	In respect of damage to fixed property
o	In respect to road traffic collisions
o	Other gratuities
(7)	Stokvel
(8)	Other income
o	Net withdrawals from savings (i.e., total withdrawals minus total deposits)
o	Non-refundable bursaries from all sources
o	Benefits, donations and gifts received from private persons, welfare funds, clubs, government, etc.
o	Cash (including bonuses from buying associations)
o	Value of food received
o	Value of housing
o	Value of clothing
o	Value of other benefits, donations, gifts
(9)	Lobola/Dowry received
(10)	All other income (e.g., from gambling, lotto winnings)
(11)	Income not elsewhere specified

As far as the capturing of income and expenditure in IES 2005/2006 is concerned, the diary method was introduced for the first time, and the respondents needed to fill in the main questionnaire as well as four weekly diaries. The main questionnaire was divided into five parts (Table 2.12). The fieldworkers visited each household five times and asked the household members to answer questions from each part in each visit so as to avoid interviewee fatigue, which might have happened in the previous IESs when the households were required to answer all questions in a single visit by the fieldworkers.

Table 2.12: Questions asked from the main questionnaire in each interview, IES 2005/2006

Interview	**Section**	
First	Section 1:	Particulars of each person in the household
	Section 2:	Area of purchase of goods by this household
Second	Section 3:	Information regarding dwellings
	Section 4:	Housing
	Section 5:	Swimming pool and garden
	Section 6:	Expenditure when away from home
	Section 7:	Domestic workers
	Section 8:	Input costs for home production
Third	Section 9:	Clothing and footwear
	Section 10:	Household textiles
	Section 11:	Furniture and equipment
	Section 12:	Recreation, entertainment and sport
Fourth	Section 13:	Education and training
	Section 14:	Reading material and stationery
	Section 15:	Health services and medical requisites
	Section 16:	Transport
	Section 17:	Computer and telecommunication equipment
Fifth	Section 18:	Finance charges, income tax and investment
	Section 19:	Particulars of income (Regular income and irregular income)

The main questionnaire asked the respondents to declare regular and irregular income as well as expenditure on semi-durable items, durable items and services, either during the month prior to the survey month, during the 11 months prior to the survey month, and/or during the past 12 months.

In addition to the main questionnaire, households were required to record their expenditures for four weeks in the form of four weekly diaries. The fieldworkers collected each diary on a weekly basis. Some households, for various reasons like moving from selected dwelling units, fatigue, did not complete all four diaries. Stats SA decided that only households that completed the main questionnaire and at least two weekly diaries were accepted (Table 2.13). Missing acquisitions for households with two or three diaries were imputed, and the imputations were done as follows (Stats SA 2006d):

- If a household had two completed diaries, expenditure from the two diaries was added together and the sum was divided by two. This average figure was then used to impute for the remaining two non-completed/missing diaries.
- If a household had three completed diaries, expenditure from the three diaries was added together and the sum was divided by three. This average figure was then used to impute for the remaining non-completed/missing diary.

Table 2.13: Inclusion of households for the final IES 2005/2006 data

Description	No. of households	Decision
Completed all diaries but did not complete the main questionnaire	325	Rejected
Completed all diaries and main questionnaire	20 960	Accepted
Non-contact	199	Rejected
Refused	480	Rejected
Completed at least two diaries and main questionnaire	184	Accepted
No usable information	14	Rejected
Vacant dwelling	1 577	Rejected
Listing error	270	Rejected
Other	728	Rejected
Completed main questionnaire but only zero or one diary	455	Rejected
Total	25 192	
Final sample size	21 144 (20 960 + 184)	

Source: Stats SA (2006d).

Next, the annualized expenditure and income figures from both the main questionnaire and the weekly diaries were calculated using the method shown in Table 2.14.

Table 2.14: Derivation of annual household income and expenditure, IES 2005/2006

Type of data item	Reference period		Annualized figure
	[A]: Diary (Survey month)	[B]: Main questionnaire	
Non-durable items	1 month	–	[A] × 12
Semi-durable items	1 month	11 months	[A] + [B]
Durable items	1 month	11 months	[A] + [B]
Services	–	1 or 12 months	[B] (if reference period is 1 month) [B] × 12 (if reference period is 12 months)
Regular income	–	1 and 11 months[#]	Monthly figure + 11-month figure[#]
Irregular income	–	12 months	[B]

[#] In IES 2005/2006, respondents were asked to declare income for the previous month and income for the 11 months prior to the survey month for all regular income items. These two figures were then added before the annualized figure was derived.

Finally, all annualized income and expenditure items from both the main questionnaire and the weekly diaries were re-categorized using the Classification of Individual Consumption According to Purpose (COICOP) method. COICOP, adopted for the first time in South Africa in IES 2005/2006, is a reference classification published by the United Nations Statistics Division that divides the purpose of individual consumption expenditures incurred by the following three institutional sectors: households, non-profit institutions serving households and general government. Table 2.15 shows that there are 11 main groups in the COICOP[14], and only the items from group 1 (i.e., CPI consumption) were included for the compilation of the revised CPI by the South African Reserve Bank (SARB)[15].

Table 2.15: The main categories of the COICOP, IES 2005/2006

Group 1: CPI Consumption (i.e., items included for the compilation of the CPI)		
(A)	Food and non-alcoholic beverages	
o	Food	
	–	Bread and cereals
	–	Meat
	–	Fish
	–	Milk, cheese and eggs
	–	Oils and fats
	–	Fruits
	–	Vegetables
	–	Sugar, jam, honey, chocolate and confectionery
	–	Food products not elsewhere classified
o	Non-alcoholic beverages	
	–	Coffee, tea and cocoa
	–	Mineral waters, soft drinks, fruit and vegetable juices
o	Unclassified expenditure on food from the diary	
(B)	Alcoholic beverages, tobacco and narcotics	
o	Alcoholic beverages	
o	Tobacco	

[14] Yu (2008: 22-32) provides more detail by showing the income and consumption items in each main group of the COICOP.

[15] This will be discussed in more detail in Chapter 5, when the CPIs are used to derive real income and expenditure.

Table 2.15: Continued

(C)	Clothing and footwear
o	Clothing
o	Footwear
(D)	Housing, water, electricity, gas and other fuels
o	Actual rentals for housing
o	Imputed rentals for housing
o	Maintenance and repair of the dwelling
o	Water supply and miscellaneous services relating to the dwelling
o	Electricity, gas and other fuels
(E)	Furnishings, household equipment and routine maintenance of the house
o	Furniture and furnishings, carpets and other floor covering
o	Household textiles
o	Household appliances
o	Glassware, tableware and household utensils
o	Tools and equipment for house and garden
o	Goods and services for routine household maintenance
(F)	Health
o	Medical products, appliances and equipment
o	Out-patient services
o	Hospital services
(G)	Transport
o	Purchase of vehicles
o	Operation of personal transport equipment
o	Transport services
o	Operational values of other modes of transport
(H)	Communication
o	Postal services
o	Telephone and telefax equipment
o	Telephone and telefax services
(I)	Recreation and culture
o	Audio-visual, photographic and information processing equipment
o	Other major durables for recreation and culture
o	Other recreational items and equipment, garden and pets
o	Recreational and cultural services
o	Newspapers, books and stationery
o	Package holidays
(J)	Education
o	Pre-primary and primary education
o	Secondary education
o	Tertiary education
o	Education not definable by level
(K)	Restaurants and hotels
o	Catering services
o	Accommodation services
(L)	Miscellaneous goods and services
o	Personal care
o	Personal effects
o	Social protection
o	Insurance
o	Financial services not elsewhere classified
o	Other services not elsewhere classified
(M)	Other unclassified expenditure
Group 2: In-kind consumption	
Group 3: Income	
Group 4: In-kind income	
Group 5: Savings	

Table 2.15: Continued

Group 6: Taxes
Group 7: Transfer to others
Group 8: Debts
Group 9: Loss
Group 10: Not CPI consumption (i.e., items not included for the compilation of CPI)
Group 11: Products not in income (i.e., income items that are not included in group 3)

As far as the income section (i.e., group 3) is concerned, the items included in this group in IES 2005/2006 are almost exactly the same as those included under the income section in IES 1995 and IES 2000, except that the following changes happened in the former:

- Two new items, 'tax refunds received' and 'imputed rent on owned dwelling 7% per year of value of dwelling', were included. The former was originally an expenditure item in IES 1995 and IES 2000 (Table 2.10). On the other hand, imputed rent, which stands for the estimated value of the use of owner-occupied dwellings, was asked for the first time in IES 2005/2006. On the main questionnaire, the households that owned dwelling at the time of the survey were asked to declare the imputed rent[16]. As Stats SA was worried that the respondents might have given inaccurate answers, rental yields, based on the value of the property provided by the respondents, were assessed. This resulted in an annual rental value of 7% of the value of the property[17]. This imputed rent item was also regarded as a consumption item under "(D) Housing, water, electricity, gas and other fuels" (See Table 2.15).
- The item 'net withdrawals from savings', which was included as in irregular income item in IES 1995 and IES 2000, was no longer included as in income item in IES 2005/2006, but rather as savings item (i.e., group 5) in the latter. On the other hand, the three irregular income items – 'value of goods and services received by virtue of occupation', 'non-refundable bursaries', and 'value of housing' – were included as irregular income items in IES 1995 and IES 2000, but were excluded (from group 3) in IES 2005/2006. Instead, these three items were included in the products not in income group (group 11). The same thing happened to the irregular income item 'value of transport', which was asked for the first time in IES 2005/2006.

In addition to the two imputed rent items mentioned above, expenditure on a few items were also asked for the first time in IES 2005/2006, such as the cost of other dwelling, and

[16] The question was asked as follows (Question 4.8 of Section 4): 'If you were to rent this dwelling, how much would you pay for it per month?'

[17] The variable 'Imputed rent 7% per year of value of dwelling' was included both as an income item (under group 3: Income) and an expenditure item (under sub-group D: housing, water, electricity, gas and other fuels in group 1: Consumption), while 'Imputed rent' was included as an item under group 10: Not CPI consumption.

expenditure on alternative means of transport like horses and donkeys. With regard to the categorization of the expenditure items, it is obvious that the STC approach used in IES 1995 and IES 2000 (i.e., the items from the first twenty expenditure categories were added to derive the total household annual expenditure) is not directly comparable with the COICOP method in IES 2005/2006. In the latter survey, the expenditure items fell under group 1 (Consumption), group 2 (In-kind consumption), group 5 (Savings), group 6 (Taxes), group 7 (Transfers to others) and group 9 (Loss).

Since the COICOP is very different from the STC in the previous IESs, in order for meaningful comparative analysis on poverty and inequality to be conducted across the three IESs, there are two options:

- Re-categorize the income and expenditure items in the 1995 and 2000 surveys, using the 2005 COICOP structure.
- Re-categorize the income and expenditure items in the 2005/2006 survey using the STC.

Both approaches will be adopted when the poverty and inequality trends are analysed in Chapter 6. Finally, regardless of whether the STC or COICOP approach was adopted, all households had specified income/consumption/expenditure in all three surveys. No households reported zero income/consumption/expenditure amounts in IES 1995, while only a very negligible proportion of households (less than 1% in each survey) had zero income/consumption/expenditure in the other two IESs.

To conclude, income and expenditure information was in actual amounts collected by the recall method in all three IESs, with the addition of the diary method in IES 2005/2006. Besides, the possible interviewer and interviewee fatigue was taken into consideration by Stats SA in the latter survey by means of multiple visits to the households to capture the required information. Finally, the method to categorise the income and expenditure in this survey was different. How these factors affect the reliability of the income and expenditure information as well as the subsequent poverty and inequality estimates will be examined in greater detail from Chapter 3 onwards.

2.4 October Household Surveys (OHSs), Labour Force Surveys (LFSs), and Quarterly Labour Force Surveys (QLFSs)

Stats SA has been collecting labour market data since 1993 with the OHS, which was conducted annually between 1993 and 1999[18], and the LFS, which was a biannual survey (conducted in March[19] and September) introduced in 2000 to replace the OHS. The QLFS was introduced in 2008 to replace the LFS and takes place four times a year[20]. Although the main objective of these surveys is to capture the labour market status of the individuals, questions relating to total income and expenditure were asked in some surveys.

2.4.1 Sampling design and sample size

As the metadata document was not provided in the OHS 1994-1995 data, the sampling methodology is not known[21]. It was only known that OHS 1994-1995 data was weighted using the 1991 Census weights, but OHS 1995 was later re-weighted using the 1996 Census weights. In OHS 1996, a sample of 1 600 Enumerator Areas (EAs) was produced in conjunction with the sample for Census 1996. The first stratification was done by province and area type. In the second phase, 10 households were drawn from each EA (Stats SA 1996). This means the sample size was approximately 16 000 households. A similar two-stage sampling procedure was adopted in OHS 1997, with the only difference being that a sample of 3 000 (instead of 1 600) EAs was produced in the first stage (Stats SA 1997a). Such two-stage sampling procedure was adopted again in both OHS 1998 and OHS 1999 (Stats SA 1998a & 1999). However, in the first stage, a sample of 2 000 EAs was produced in the first stage in OHS 1998, while the number of EAs amounted to 3 000 in OHS 1999, before 10 households were selected from each EA in the second stage.

As the LFS 2000a was considered to be a pilot study for the newly introduced LFS which replaced the OHS, the sample size was much smaller (less than 10 000 households). The households were drawn in 1 574 enumerator areas (EAs), with 10 households drawn in each of the 426 non-urban EAs and five households drawn in each of the 1 149 urban EAs. The two-stage sampling methodology as discussed above was adopted. Finally, OHS 1996 to LFS

[18] The 1996 OHS actually took place in November as enumeration for Census 1996 was conducted during that time.
[19] The only exception is the first round of the 2000 LFS, which took place in February.
[20] For the remainder of the dissertation, the OHSs will be referred to as OHS 1993, OHS 1994, etc., while the LFSs will be referred to as LFS 2000a (for the first round of LFS in 2000), LFS 2000b (second round in 2000), LFS 2001a, and so forth. Finally, the QLFSs will be referred to as QLFS 2008Q1 (for the first round of QLFS in 2008), QLFS 2008Q2 (second round in 2008), and so forth.
[21] The author contacted Stats SA for the 1993-1995 metadata documents, but did not receive such information.

2000a were weighted using the 2001 Census weights (Stats SA 2000a).

A different two-step sampling procedure was introduced in the LFS 2000b and it was adopted until the LFS 2004a (Stats SA 2000a, 2000b, 2001a, 2001b, 2002a, 2002b, 2003a, 2003b & 2004a). The first stage was the selection of about 3 000 Primary Sampling Units (PSUs), with probability proportional to size of PSU_S, from the 2001 Census list of Enumerator Areas (EAs) to form the Master Sample. The second stage involved the systematic selection of 10 dwelling units from each of the selected PSUs. The Master Sample was stratified into eighteen strata, i.e. nine provinces and within each province by urban / non-urban. In addition to the adoption of this new sampling procedure, a rotating panel approach was adopted from LFS 2001b until LFS 2007b, as the selected dwelling units would remain in the sample for five consecutive surveys (e.g., a duration of two and a half years), with 20% of these dwelling units rotating out at each round of the survey[22].

A similar two-stage sampling methodology was adopted between the 2004 September and 2007 September LFSs (Stats SA 2004b, 2005a, 2005b, 2006a, 2006b, 2007a & 2007b). In the first stage, EAs with a household count of fewer than 25 were excluded from the Census 2001 frame that was used to draw the sample of 3 000 PSUs for the Master Sample. Moreover, EAs in the census database that were found to have fewer than 60 dwelling units during listing were pooled. Next, the 3 000 PSUs were allocated to the 53 DCs using the power allocation method. The PSUs was then sampled using probability proportional to size principles, with the measure of size used being the number of households in a PSU as calculated in the census.

This sampling methodology was adopted again with the introduction of the QLFS (Stats SA 2008a). In addition, the rotating panel approach was also adopted, but this time the selected dwelling units would remain in the sample for four consecutive surveys (i.e., a duration of one year), with 25% of these dwelling units rotating out at each round of the survey. Finally, LFS 2004b to QLFS 2009Q4 were also weighted with Census 2001 weights.

Table 2.16 reports the number of sampled households in OHSs/LFSs/QLFSs. As these surveys will be used to derive labour market trends in Chapter 4, the number of sampled individuals of all ages and the working-age population (i.e., people aged 15-65 years) in the surveys under investigation are also presented in the table. With the exception of OHS 1996

[22] The advantage of this type of design is that it "provides the basis for monitoring changes in the work situation of members of the same households over time, while retaining the larger picture of the overall employment situation in the country. It also allows for both longitudinal and cross-sectional analysis" (Stats SA 2004a).

(which coincided with the 1996 Census), OHS 1998 (for which funding restrictions were more severe) and LFS 2000a (which was considered to be a pilot study for the newly introduced LFSs), all surveys consisted of samples of around 25 000 – 28 000 households.

Table 2.16: Sample size in each OHS, LFS and QLFS

Survey	Number of households	Sample size – All ages	Sample size – 15-65 years
OHS 1994	30 279	132 469	82 446
OHS 1995	29 700	130 787	81 108
OHS 1996	15 920	72 889	44 001
OHS 1997	29 811	140 015	82 613[23]
OHS 1998	18 968	82 213[24]	49 560[24]
OHS 1999	26 134	106 650	65 995
LFS 2000a	9 705	38 529	23 713[25]
LFS 2000b	26 648	105 370	65 612[26]
LFS 2001a	28 170	107 726	67 903
LFS 2001b	27 356	106 439	66 517[27]
LFS 2002a	29 010	109 408	69 150
LFS 2002b	26 474	102 480	64 372
LFS 2003a	26 702	100 834	63 825
LFS 2003b	26 825	98 748	62 869
LFS 2004a	26 829	98 256	62 696
LFS 2004b	28 594	109 888	68 433[28]
LFS 2005a	28 841	110 671	69 101
LFS 2005b	28 418	109 079	68 269
LFS 2006a	28 649	108 345	68 386
LFS 2006b	28 363	106 900	66 867
LFS 2007a	29 466	109 551	68 673
LFS 2007b	27 981	105 986	65 891
QLFS 2008Q1	26 180	95 186	59 488
QLFS 2008Q2	26 293	93 945	58 540
QLFS 2008Q3	26 619	93 725	58 315
QLFS 2008Q4	26 817	93 062	57 944
QLFS 2009Q1	26 655	92 777	57 726
QLFS 2009Q2	25 885	90 783	56 473
QLFS 2009Q3	25 163	88 252	54 957
QLFS 2009Q4	24 917	87 653	54 893

[23] In the person file, 82 613 people were aged between 15 and 65 years, but 6 of them were absent in the worker file.

[24] Although there were 82 263 observations in the person file, 50 of them existed twice in the dataset (of these 50 people, 37 of them fell under the working-age population). Therefore, the correct sample size should be 82 213 (82 263 – 50) for all ages and 49 560 (49 597 – 37) for the working age population. However, looking at the worker file, there were 49 599 people aged between 15 and 65 years. It was found that 39 people (49 599 – 49 560) were present in the worker file absent in the person file. All 39 people who went missing in the person file come from Western Cape. 36 of them are Coloureds and the remaining 3 are Blacks. These 39 people will be excluded from all the OHS 1998 analyses for the remainder of the dissertation.

[25] In the person file, there were 23 713 people aged 15-65 years, but 13 of them did not exist in the worker file.

[26] In the person file, there were 65 612 people aged 15-65 years, but 113 of them did not exist in the worker file.

[27] In the person file, there were 66 517 people aged 15-65 years, but 1 of them did not exist in the worker file.

[28] In the person file, there were 68 433 people aged 15-65 years, but 1 of them did not exist in the worker file.

2.4.2 The questionnaire

2.4.2.1 Questionnaire structure

In general, each of the surveys under investigation consisted of the following sections:

- Section 1: Particulars of each person in the household were asked.
- Section 2: A few important questions covering economic activities were asked, which in turn determined the labour market status (i.e., employed, unemployed, inactive) of the individuals.
- Section 3: This section asked questions relating to unemployment and non-economic activities, such as how long the unemployed had been looking for work, reasons for the inactive not seeking work, etc.
- Section 4: Only the employed were asked to take part. Questions were asked about the work situation of the employed, such as occupation, industry and weekly work hours.

Until LFS 2005a, there was a section (Section 7 in most surveys) asking questions regarding the households (e.g., dwelling type, water source, etc.). However, these questions were not asked again from LFS 2005b. It was because the main objective of the OHS/LFS/QLFS is to capture the information on the labour market, and so the household-level questions (which included the household total expenditure question) were gradually excluded from these surveys, and were asked in the GHSs instead. In addition, in selected OHSs and LFSs, some sections (Section 5 and 6 in most surveys) were devoted to ask questions relating to the respondents' participation status in the expanded public works program (EPWP), agricultural and uncompensated activities in the past 12 months, information on migrant workers, or death of household members in the past 12 months. However, these questions were not asked anymore since the inception of the QLFS.

2.4.2.2 Labour market status questions

As mentioned in Section 2.4.1, as the metadata document was not provided in the OHS 1994-1995 data, the labour market status derivation methodology is not known. Looking at the post-1995 data, questions from Sections 2 and 3 of the questionnaire were used to derive both the narrow (strict) and broad (expanded) labour market status of the respondent.

Two standard definitions of unemployment were utilized by Stats SA, namely the narrow definition and broad definition of unemployment. Individuals were generally defined as narrowly unemployed if they: (a) did not work during the seven days prior to the interview, (b) wanted to work and would accept a job if being offered one (there was an additional requirement since LFS 2000b that these people must be available to start work within two weeks of the interview), and (c) had taken active steps to look for work or to start a business in the four weeks prior to the interview. Those who only met the first two requirements above were defined as discouraged job seekers, and were classified as inactive under the narrow definition but unemployed under the broad definition (Table 2.17).

Table 2.17: Derivation of narrow and broad labour force participation rates and unemployment rates, OHSs/LFSs/QLFSs

Labour market status (1) Employed (2) Unemployed (3) Discouraged job seeker[#] (4) Inactive
Narrow labour force participation rate = Labour force[##] / Working-age population[###] $= \frac{(1)+(2)}{(1)+(2)+(3)+(4)}$
Broad labour force participation rate = Labour force / Working-age population $= \frac{(1)+(2)+(3)}{(1)+(2)+(3)+(4)}$
Narrow unemployment rate = Unemployed / Labour force $= \frac{(2)}{(1)+(2)}$
Broad unemployment rate = Unemployed / Labour force $= \frac{(2)+(3)}{(1)+(2)+(3)}$

[#] These people were defined as inactive and unemployed under the narrow and broad definitions respectively, in the OHS/LFS labour market status derivation methodologies.
[##] Labour force (LF), also known as economically active population (EAP), stands for the total number of people in the working ages (15-65 years) who are willing and able to work.
[###] Working-age population stands for people aged between 15 and 65 years.

However, looking at the methodology to derive labour market status from OHS 1995 to QLFS 2009Q4, there have been numerous changes throughout the years. In addition, the broad methodologies before and after the introduction of QLFS are very different. These changes could result in the incomparability of the labour market trends across the surveys. Chapter 3 will come back to this issue in greater detail when examining the factors affecting the comparability and reliability of labour market trends, and Chapter 4 will address these issues, if possible, before re-visiting the trends in the labour market aggregates.

2.4.2.3 Household income and expenditure questions

In the OHSs, total household income was only captured in 1999[29]. As far as the total expenditure is concerned, the question was asked in OHS 1996-1999. Between 1996 and 1998, the household was asked to declare total expenditure in actual amounts during the past month, while in 1999, the household was asked to declare the relevant monthly expenditure category[30]. The household expenditure question with the same categorical answers as in OHS 1999 was asked again in LFSs, but only in four surveys (i.e., the 2001-2004 September surveys)[31]. In the QLFSs, both household income and expenditure questions were not asked. One problem of the OHS/LFS expenditure data is that there were too few expenditure categories, and a high proportion of households (about two-thirds) fell into the first three categories (See Table 2.18). Interestingly, this is not the case when looking at household income in OHS 1999, as only 49.2% of households fell into the first three categories.

Table 2.18: Proportion of households in each monthly household income or expenditure category, OHSs/LFSs

	OHS				
	1996#	**1997#**	**1998#**	**1999##**	**1999## (Income)**
R0 – R399	21.5%	23.3%	26.0%	25.0%	15.9%
R400 – R799	24.6%	29.1%	29.2%	26.3%	20.9%
R800 – R1 199	13.8%	14.8%	12.6%	13.2%	12.4%
R1 200 – R1 799	9.0%	8.8%	7.9%	8.5%	11.4%
R1 800 – R2 499	7.1%	5.5%	6.1%	5.8%	7.0%
R2 500 – R4 999	11.1%	8.7%	9.1%	7.8%	11.0%
R5 000 – R9 999	4.1%	3.8%	3.8%	4.3%	7.2%
R10 000 or more	1.0%	1.1%	0.9%	1.6%	5.0%
Don't know / Refuse / Unspecified	***7.6%***	***5.0%***	***4.5%***	***7.6%***	***9.3%***
	100.0%	100.0%	100.0%	100.0%	100.0%
R0 – R1 199	59.9%	67.2%	67.8%	64.5%	49.2%

The household expenditure was declared as actual amounts in OHS 1996-1998, and answers were used to categorize respondents into the expenditure categories in the table.

The household weight in OHS 1999 was inaccurate, as the weighted number of households was seriously under-estimated, and the racial shares of the households were incorrect (Yu 2007: 38-39). Hence, the average of the person weights of the members of the household was used instead as a proxy for household weight.

---

[29] The total household income question was asked as follows in OHS 1999 (Question 6.36, Section 6): "What was the total household income in the last month, including wage, salaries, government grants, private pensions and all other sources of income?" The respondent then could choose from the following 10 income categories: 1: R0 – R399, 2: R400 – R799, 3: R800 – R1 199, 4: R1 200 – R1 799, 5: R1 800 – R2 499, 6: R2 500 – R4 999, 7: R5 000 – R9 999, R10 000 or more, 9: Don't know, 10: Refuse.

[30] The question was asked as follows in each OHS 1996 (Question 1.39, Section 1), OHS 1997 and OHS 1998 (Question 9.40, Section 9): "How much money did this household spend in total, on all items (including food, clothing, housing, transport, medical care, etc), during the past month?" On the other hand, the question was asked as follows in OHS 1999 (Question 6.31, Section 6): "What was the total household expenditure in the last month? Include everything that the household and its members spent money on, including food, clothing, transport, rent and rates, alcohol and tobacco, school fees, entertainment and any other expenses." There were 10 categories for the respondents to choose from: 1: R0 – R399, 2: R400 – R799, 3: R800 – R1 199, 4: R1 200 – R1 799, 5: R1 800 – R2 499, 6: R2 500 – R4 999, 7: R5 000 – R9 999, R10 000 or more, 9: Don't know, 10: Refuse.

[31] The household expenditure questions in these four September LFSs (Question 6.25, Section 6 in LFS 2001; Question 7.25, Section 7 in LFS 2002; Question 7.29, Section 7 in LFS 2003; and Question 7.30, Section 7 in LFS 2004) were asked in exactly the same way as in OHS 1999.

Table 2.18: Continued

	LFS (September)			
	2001	2002	2003	2004
R0 – R399	31.7%	30.6%	24.6%	21.8%
R400 – R799	26.1%	26.3%	28.8%	28.6%
R800 – R1 199	11.6%	12.1%	13.9%	14.1%
R1 200 – R1 799	7.2%	7.4%	7.6%	8.4%
R1 800 – R2 499	5.4%	5.8%	5.9%	6.6%
R2 500 – R4 999	8.1%	7.5%	8.4%	8.3%
R5 000 – R9 999	4.6%	5.2%	5.7%	6.6%
R10 000 or more	1.7%	2.2%	2.9%	3.1%
Don't know / Refuse / Unspecified	***3.7%***	***3.0%***	***2.3%***	***2.7%***
	100.0%	100.0%	100.0%	100.0%
R0 – R1 199	69.4%	68.9%	67.3%	64.4%

In the 1996-1998 OHSs, no households reported zero income or expenditure. In addition, since the lowest expenditure category was "R0 – R399", it is not possible to distinguish the zero-expenditure households (if any). As far as the households with unspecified income or expenditure are concerned, the proportion of households falling into such category was not low (such proportion was nearly 8% in OHS 1996 and above 9% in OHS 1999).

Comparing the respondents' answers in the household income and expenditure questions in OHS 1999, the results on Table 2.19 suggest that the households tended to under-declare their household expenditure, compared with their answers on household income. For example, looking at households in the R800 – R1 199 expenditure category, only 36.2% of them fell into the R800 – R1 199 income category. However, 6.4% (1.8% + 4.6%) reported that the household income was below R800, while the income of 53.2% (22.4% + 12.7% + 13.2% + 3.7% + 1.2%) of the households exceeded R1 199.

Table 2.19: Household income vs. Household expenditure, OHS 1999

		Household expenditure category								
		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Household income category	[1]	56.0%	4.4%	1.8%	0.8%	0.7%	0.3%	0.3%	0.1%	5.2%
	[2]	25.5%	50.2%	4.6%	2.1%	0.9%	0.6%	0.5%	0.0%	5.2%
	[3]	7.2%	19.2%	36.2%	3.4%	1.7%	0.6%	0.5%	0.4%	4.0%
	[4]	5.3%	12.0%	22.4%	38.4%	5.4%	1.6%	0.2%	0.3%	2.7%
	[5]	1.7%	5.7%	12.7%	16.7%	26.7%	3.2%	0.9%	1.1%	1.7%
	[6]	1.0%	4.2%	13.2%	24.3%	38.6%	39.9%	5.1%	2.9%	2.9%
	[7]	0.2%	0.8%	3.7%	7.6%	16.5%	35.8%	42.6%	3.8%	2.3%
	[8]	0.1%	0.3%	1.2%	2.2%	4.2%	13.3%	43.4%	84.3%	2.3%
	[9]	3.1%	3.2%	4.3%	4.5%	5.4%	4.8%	6.7%	7.1%	73.7%
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Note: Income or expenditure categories: [1]: R0 – R399, [2]: R400 – R799, [3]: R800 – R1 199, [4]: R1 200 – R1 799, [5]: R1 800 – R2 499, [6]: R2 500 – R4 999, [7]: R5 000 – R9 999, [8]: R10 000 or more, [9]: Don't know / Refuse / Unspecified

2.5 General Household Surveys (GHSs)

The GHS, which is also conducted by Stats SA, was introduced in 2002. The two main objectives of the survey are to measure the multiple facets of the living conditions of South African households, as well as the quality of service delivery in a number of key service sectors. Six broad areas are covered by the GHS, namely education, health, activities related to work and unemployment, non-remunerated trips undertaken by the household, housing, and household access to services and facilities.

2.5.1 Sampling design and sample size

Between 24 000 and 28 000 households took part in each GHS[32]. The sampling methodologies adopted in GHS 2002-2005 and GHS 2006-2007 were exactly the same as the approach used in LFS 2000b-2004a and LFS 2004b-2007b respectively (refer to Section 2.4.1) (Stats SA 2003d, 2004c, 2005c, 2006d, 2007c & 2008g).

In GHS 2008 and 2009, the master sample used a two-stage, stratified design with probability-proportional-to-size (PPS) sampling of PSUs from within strata, and systematic sampling of dwelling units (DUs) from the sampled primary sampling units (PSUs). A self-weighting design at provincial level was used and stratification was divided into two levels. Primary stratification was defined by metropolitan and non-metropolitan geographic area type. During secondary stratification, the Census 2001 data were summarised at PSU level, with the household size, education, occupancy status, gender, industry and income variables being used. Next, a Randomised Probability Proportional to Size (RPPS) systematic sample of PSUs was drawn in each stratum, with the measure of size being the number of households in the PSU. About 3 080 PSUs were selected altogether. In each selected PSU a systematic sample of dwelling units was drawn. The number of DUs selected per PSU varied from PSU to PSU, depending on the Inverse Sampling Ratios (ISR) of each PSU (Stats SA 2009 & 2010).

2.5.2 The questionnaire

2.5.2.1 Questionnaire design

In general, the GHS consists of the following four sections:

- Section 1: Details of each person in the household were asked, and the questions are

[32] The sample size at household level in each GHS is as follows: 26 243 in 2002, 26 398 in 2003, 26 214 in 2004, 28 129 in 2005, 28 002 in 2006, 29 280 in 2007, 24 222 in 2008 and 25 303 in 2009.

very similar to those asked in the LFSs and QLFSs, with the addition that questions relating to health, disability and receipt of social grants were also asked.

- Section 2: A few important questions covering economic activities were asked, which in turn determined the labour market status (i.e., employed, unemployed, inactive) of the individuals. In addition, some questions relating to the working conditions of the employed were asked. Since the aim of the GHS is not to capture labour market information, the labour market status questions asked in Section 2 are fewer and less detailed than the questions asked in Sections 2-4 of the LFSs and QLFSs.
- Section 3: This section collected information about trips taken by one or more household members in the past 12 months. The only exception is GHS 2002, which collected information regarding children ever born instead.
- Section 4: In this section, household-level questions were asked, such as dwelling type, water source, ownership of household goods and services, etc. In addition, questions relating to household expenditure as well as child and adult hunger in the past 12 months were also asked.

2.5.2.2 Labour market status questions

GHS 2002-2007 adopted the methodology used in LFS 2000b-2007b (to be discussed in Chapter 3) to derive the labour market status, while the GHS 2008-2009 data did not provide any labour market status variables. The metadata documents in both datasets did not explain the reason why the variable was excluded, and whether the QLFS labour market status methodology would be adopted in the forthcoming GHSs.

2.5.2.3 Household income and expenditure questions

Household income was not captured by the GHS, except that GHS 2009 only derived the income of households with total monthly income of less than R20 000 in nominal terms by adding the incomes of members from earnings, social grants and remittances. On the other hand, the households were asked to declare the relevant monthly household expenditure category[33], and the categories were exactly the same as in the LFSs in all GHSs, except that in GHS 2009, an improvement was made in providing a more detailed break-down of the households with R0-R399 expenditure, with the introduction of the categories 'R0', 'R1-R299' and 'R200-399' to replace the original 'R0-R399' category. Hence, it is possible to

[33] The expenditure question (Question 4.45 in GHS 2002, Question 4.63 in GHS 2003, Question 4.71 in GHS 2004, Question 4.79 in GHS 2005, Question 4.69 in GHS 2006-2008 and Question 4.20 in GHS 2009) was asked as "What was the total household expenditure in the last month? Include everything that the household and its members spent money on, including food, clothing, transport, rent and rates, alcohol and tobacco, school fees, entertainment and any other expenses."

find out the proportion of households with zero expenditure.

Table 2.20 shows that a similar problem occurs in the GHSs as in the OHSs and LFSs, namely the very high proportion of households reporting a monthly expenditure of less than R1 199. However, such proportion shows a continuous downward trend (from 70.4% in 2002 to 47.6% in 2009), and this could be due to the fact that households would move to higher nominal expenditure categories as time goes by due to the impact of inflation. Furthermore, compared with OHSs/LFSs, the proportion of households with unspecified expenditure in the GHSs was lower.

Table 2.20: Proportion of households in each monthly household expenditure category, GHSs

	2002	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
R0	N/A							0.5%
R1 – R199								2.4%
R200 – R399								8.4%
R0 – R399	31.3%	26.0%	18.5%	19.2%	17.6%	13.5%	9.4%	11.3%
R400 – R799	27.2%	27.6%	28.6%	28.0%	28.9%	27.2%	23.1%	19.5%
R800 – R1 199	11.9%	13.7%	14.1%	15.0%	17.5%	17.9%	19.1%	16.8%
R1 200 – R1 799	7.1%	7.6%	10.4%	10.3%	10.5%	11.9%	12.4%	12.3%
R1 800 – R2 499	5.5%	5.9%	6.7%	6.4%	6.5%	7.1%	8.6%	8.9%
R2 500 – R4 999	7.2%	7.9%	10.0%	10.3%	9.2%	10.7%	11.5%	10.8%
R5 000 – R9 999	4.7%	5.3%	6.4%	6.3%	6.0%	7.3%	8.0%	9.3%
R10 000 or more	1.7%	2.3%	2.2%	2.5%	2.7%	3.0%	5.3%	6.7%
Don't know / Refuse / Unspecified	***3.5%***	***3.7%***	***3.1%***	***2.1%***	***1.4%***	***1.5%***	***2.6%***	***4.4%***
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
R0 – R1 199	70.4%	67.2%	61.2%	62.1%	63.9%	58.6%	51.6%	47.6%

The results in Tables 2.18 and 2.20 show that with the exception of GHS 2009, the number of expenditure intervals and the width of each interval are exactly the same in OHSs, LFSs and GHSs. The intervals are fewer than in the two censuses and CS 2007, but the width of each interval is narrower. In addition, a high proportion of households reporting spending less than R1 200 per month in nominal terms in these surveys. Furthermore, the proportion of households with unspecified expenditure in the GHSs was lower than what happened in thw two censuses and CS 2007. How these issues affect the poverty and inequality estimates and trends will be discussed in Chapters 3 and 6.

2.6 The Project for Statistics on Living Standards and Development (PSLSD)

The PSLSD survey was conducted in 1993 by SALDRU with assistance from the World Bank. The survey collected a wide range of indicators of standard of living, with the main aim being the collection of statistical information about the conditions under which South Africans lived. The information was intended to provide policy makers with the data required for planning strategies to implement the goals outlined in the Government's Reconstruction and Development Programme (RDP). The survey only took place once and the data was released in 1994.

2.6.1 Sampling design and sample size

The sample design adopted for the study was a two-stage self-weighting design (SALDRU 1994) in which the first stage units were Census 1991 Enumerator Subdistricts (ESDs) and the second stage were households. In the first stage, the ESDs were selected with probability proportional to size, based on the census population. Systematic sampling was used throughout, i.e., sampling at fixed interval in a list of ESDs, starting at a randomly selected starting point. The primary objective was to ensure that the racial and geographic breakdown approximated the national population distribution, and this was done by listing the area stage units (ESDs) according to statistical region and then within the statistical region by area type.

In the second stage, in each selected ESD, a listing or enumeration of households was carried out by a field operation. From the households listed in an ESD, a sample of households was chosen by systematic sampling. With regard to the sample size, 8 809 households took part in the survey.

2.6.2 The questionnaire

2.6.2.1 Questionnaire structure

There was a detailed household questionnaire, with 11 sections in total, covering various topics, ranging from demographic information, employment status, income from employment and non-employment sources, food spending and non-food spending, to health and perceived quality of life. In addition, a community questionnaire was run in each cluster of the sample. Its purpose was to capture information on the availability of facilities to the community in each cluster, such as infrastructure, education, health, and recreational facilities.

2.6.2.2 Labour market status questions

Section 8.1 of the household questionnaire asked numerous questions which helped distinguishing the employed from the unemployed and inactive in narrow terms. However, the questions were not asked in too great detail, as in OHSs/LFSs/QLFSs.

2.6.2.3 Household income and expenditure questions

Household income was derived by adding regular employment income, casual or temporary work income, self-employment income and non-employment income in actual amounts (Section 8 of the questionnaire) as declared by the respondents. The recall method was used. On the other hand, household expenditure was derived by adding the respondents' answers on food spending and non-food spending (Sections 3 and 4 of the questionnaire respectively), and again, the recall method was used and the respondents declared actual amounts. In the sample, no households had zero or unspecified household income and expenditure amounts.

Looking at the derivation of household expenditure in greater detail, respondents were asked to report the spending on 31 food items, regular non-food spending on 13 items and occasional non-food spending on another 13 items on four pages of the questionnaire, as shown in Table 2.21. The degree of disaggregation is smaller than compared with the IESs (e.g., 40 pages of the IES 2000 questionnaire was devoted to capturing expenditure detail). For instance, in PSLSD, the respondents were asked to report the total spending on 'fresh milk / sour milk / yoghurt', but the IES 2000 participants were asked to declare expenditure on the following milk items: 'fresh milk', 'sour milk, maas, butter milk', 'cream', yoghurt', 'ice cream (sorbet, full cream and frozen yoghurt, etc.)', and 'milk power and whiteners'. As another example, the PSLSD participants were asked to declare the total regular non-food spending on 'cigarettes, tobacco', but the IES participants were required to provide detailed expenditure information on the following items: 'cigars and cigarillos', 'pipe and cigarette tobacco, snuff, etc.', 'other items smoked (e.g., dagga)', and 'smokers requisites, e.g., pipes, cigarette lighters, flints'.

How the level of disaggregation of items affects the accuracy of income and expenditure information and the subsequent poverty and inequality estimates will be discussed from Chapter 3 onwards.

Table 2.21: Food and non-food items included for the derivation of aggregate expenditure, PSLSD

Food spending	
1: Maize grain/samp	16: Baby Formula/Milk Powder
2: Mealie Meal/Maize Flour	17: Sugar
3: Rice	18: Mutton/Beef/Pork
4: White/Brown Bread	19: Chicken
5: Wheat Flour	20: Eggs
6: Breakfast Cereal	21: Fresh Fish
7: Dried Peas/Lentils/Beans	22: Tinned Fish
8: Potatoes	23: Pumpkin/squash
9: Tomatoes	24: Other vegetables
10: Madumbes/Sweet Potatoes/Other roots/Tubers	25: Bananas
11: Vegetable Oil	26: Apples
12: Margarine/Butter/Ghee/Other Fats	27: Citrus fruit
13: Cheese	28: Soft drinks
14: Jam	29: Meals prepared outside home
15: Fresh Milk/Sour Milk/Yoghurt	30: Meals given to guests
	31: Other food expenditure/consumption
Non-food spending	
Regular food spending	Occasional non-food spending
1: Cigarettes, tobacco	1: Kitchen equipment, like pots and pans, lamps, torches and so on
2: Beer, wine, spirits	2: Home maintenance and repairs to the dwelling
3: Entertainment (cinema, sports, music, etc)	3: Bedding, sheets, blankets and towels
4: Personalised care items: soap, shampoo, haircuts, facials, and so on	4: Furniture and other household appliances
5: Newspapers/stationery, envelopes, stamps	5: Shoes for adults and children
6: Telephone	6: Clothes for adults and children
7: Petrol, oil and car service	7: Material to make clothing or curtains
8: Buses, taxis, trains and air tickets	8: Dentists, doctors or nurses
9: Washing powder	9: Hospital fees
10: Childcare	10: Medical supplies, for example, medicines, bandages and so on
11: Religious and membership dues of organisations	11: Traditional healer's fees
12: Informal taxation and donations	12: Holidays
13: Domestic servants, gardeners and other household labour	13: Jewellery, watches, other luxury good

2.7 National Income Dynamics Study (NIDS)

The NIDS, conducted by SALDRU, is South Africa's the first national panel study of individuals of all ages, with the data from the first wave (2008) being released in the last quarter of 2009. The data was revised in November 2010, and the revised version is used in this dissertation. The second wave took place in 2010. The main objective of NIDS is reported to be measuring and understanding who is getting ahead and who is falling behind in South Africa, as well as investigating the reasons why some people are making progress and the others are not. The five broad areas covered by NIDS are as follows: income and expenditures of the household and the individuals, the assets owned by the household and the household's access to services, educational attainment level and health status, labour market status of the individuals, and membership of community groups.

2.7.1 Sampling design and sample size

A stratified, two-stage cluster sample design was adopted (SALDRU 2009). In the first stage, 400 primary sampling units (PSUs) were selected from Stats SA's 2003 Master Sample of 3000 PSUs[34]. The sample of PSUs was a subset of the Master Sample. The explicit strata in the Master Sample were the 53 District Councils (DCs). Next, eight non-overlapping samples of dwelling units were systematically drawn within each PSU. Each of these samples was called a 'cluster' by Stats SA.

Fieldworkers were instructed to interview all households living at the selected dwelling unit. If it was found that the dwelling unit was unoccupied or the dwelling no longer existed, the fieldworkers were not permitted to substitute the dwelling unit, but recorded this information on the household control sheet. In the first wave, 7 302 households took part in the survey.

2.7.2 The questionnaire

2.7.2.1 Questionnaire structure

There are four questionnaires in total. First, the household questionnaire asked questions on the particulars of each member of the household, household facilities, expenditure on food and non-food items, as well as the experience of positive events (e.g., inheritance, household member(s) receiving a scholarship) and negative events (e.g., death of family members, job loss) in the past twenty-four months. Next, respondents aged 15 years or above were asked to take part in the adult questionnaire, and were asked questions on the labour market status, income from work and non-work activities, educational attainment and health. For those adult household members who were absent at the time of the survey and were unable to take part in the adult questionnaire, a proxy questionnaire was administered to a knowledgeable household member, who in turn declared the information on the labour market status, income from work and non-work activities and educational attainment of the absent members. Finally, children aged 0-14 years were asked to take part in the child questionnaire. In both adult and child questionnaires, the respondents were asked, in a voluntary basis, to take part in a numeracy test, and each respondent was given 10 minutes to complete 15 numeracy questions, with the level of difficulty of the questions determined by the respondent's educational attainment. There were four levels of difficulty.

[34] This Master Sample was the sample used in the LFSs and GHSs between 2004 and 2007 as well as the IES 2005/2006.

2.7.2.2 Labour market status questions

As mentioned in Section 2.7.2.1, the adult and proxy questionnaire contained questions which helped capturing the labour market status of the working-age population, and the methodology adopted to derive such status was the same as in QLFSs (to be discussed in Chapter 3).

2.7.2.3 Household income and expenditure questions

Household expenditure was derived in two ways, with the recall method being adopted in both methods. First, it was derived by adding the respondents' answers on food spending, non-food spending and rent expenditure (Sections D and E of the household questionnaire), i.e., the aggregation approach Secondly, the household head was asked to declare total expenditure in the last 30 days[35], i.e., the 'one-shot' amount, single estimate approach. However, since the response rate of this question was only about 79% and SALDRU was concerned that this question would give lower total expenditure, it was decided that the total household expenditure would be derived using the first method for all households.

Table 2.22 provides more detail by showing the food and non-food items included for deriving household expenditure under the aggregation method. As in PSLSD, the extent of disaggregation of expenditure items in NIDS is smaller when compared with the IESs.

Table 2.22: Food and non-food items included for deriving household expenditure, NIDS

Food spending	
1: Mealie meal	18: Peanut butter
2: Samp	19: Milk, cheese, yoghurts and dried milk
3: Flour and bread	20: Eggs
4: Rice	21: Sugar, jam, honey, chocolates and sweets
5: Pasta	22: Soft drinks and juices
6: Biscuits, cakes, rusks	23: Tinned fruit and vegetables
7: Red meat (beef, mutton, pork, etc)	24: Breakfast cereal and porridge
8: Canned red meat	25: Baby food and baby formula
9: Chicken	26: Salt and spices
10: Fresh fish and shell fish	27: Soya products
11: Tinned fish	28: Coffee and tea
12: Dried peas, lentils, beans	29: Food hampers
13: Potatoes	30: Readymade meals brought into the 31:
14: Other vegetables	household
15: Fruits and nuts	31: Meals prepared outside the home (incl.
16: Oil for cooking	restaurants and take-aways)
17: Margarine, butter, ghee, other fats	32: Other food expenditure

[35] The question was asked as (Question D31, Section D) "How much money did this household spend on all its expenses in the last 30 days?"

Table 2.22: Continued

Non-food spending
1: Cigarettes and tobacco
2: Beer, wine and spirits
3: Entertainment such as cinema, music, MNET and DSTV
4: Sport including sporting equipment, gym and club membership
5: Personal care items such as cosmetics, soap, shampoo and haircuts
6: Jewellery and watches
7: Newspapers, stationery, envelopes, stamps and books excluding school books
8: Cell phone account and/or airtime
9: Telephone account
10: Lotto, gambling and horse-racing
11: Internet if not included in the telephone account
12: Trips and holidays excluding transport costs
13: Ceremonies such as weddings and funerals
14: Car payments excluding insurance
15: Petrol, oil and car service
16: Buses, taxis, trains and air tickets including transport to school
17: Water
18: Electricity
19: Other energy sources such as wood, paraffin, charcoal/coal, candles, gas, purchasing/charging batteries and diesel oil for generators
20: Municipal rates
21: Levies for example sectional title, share block and timeshare
22: Life insurance
23: Funeral policies or burial societies
24: Educational policies
25: Short-term insurance for example car, property & fire and crop insurance
26: Kitchen equipment, like pots and pans, cutlery and crockery
27: Home maintenance and repairs to the dwelling
28: Bedding, sheets, blankets and towels
29: Material to make curtains and other household items
30: Hire purchase (HP) payments on furniture and other household appliances
31: Furniture and other household appliances bought with cash or by credit card
32: Shoes and clothes (excluding school uniforms) bought with cash or by credit card
33: Account payments on shoes and clothes excluding school uniforms
34: Material to make clothing
35: Medical aid schemes/medical insurance such as hospital plan
36: Dentists, doctors or nurses
37: Hospital fees
38: Medical supplies, for example, medicines and bandages
39: Traditional healer's fees
40: Homeopaths, physiotherapists, dieticians
41: School fees and tuition
42: School books including stationery
43: Uniforms
44: Other school expenses such as school outings, meals at school, boarding fees, contributions to school buildings, extra costs for teachers and extramural activities
45: Washing powder, dishwashing liquid, polish and all household cleaners
46: Crèche and childcare
47: Religious and membership dues of organisations, donations to charity
48: Domestics, gardeners and other household help
49: Swimming pool maintenance
50: Pets
51: Toys
52: Gifts
53: Income tax payments

The recall method was also adopted to derive household income. The household was asked (Question D38, Section D of the household questionnaire) to declare total household income amount received in the last month[36]. 5 446 households (out of 7 302) gave specific answer to this question. For the remaining 1 856 households, they were asked (Question D39, Section D of the household questionnaire) to declare the relevant total monthly household income category and could choose from 15 categories[37]. The household income amount was then derived using the mid-point of the interval. Only 474 (out of 1 856) households answered this question. In other words, 1 382 households (7 302 – 5 446 – 474) did not give any specific answers on both questions D38 and D39.

SALDRU (2009) argued that the two questions mentioned above, which adopted the single estimation approach, could result in lower household income, and opted to use the respondents' answers on each income component in Section E and Section F of the adult questionnaire (i.e., the aggregation approach). In these two sections, questions on employment income and non-employment were asked, with the respondents being asked to declare the actual amounts, and again, the recall method was used. The household income amount was then derived by adding the respondents' answers on the following seven broad components: wage income, government grant income, other government income, investment income, remittances income, implied rent income and agricultural income. This method worked successfully in 7 103 households, but the remaining 199 households did not give clear answers in the first five income components mentioned above. Therefore, in these households, the household income amount was derived by either adding their 'one-shot' income amount derived from question D38 and the implied rent income together (185 households in total) or adding the mid-point of the reported income band from Question D39 and the implied rent income together (14 households in total).

The income and expenditure variables derived by the aggregation approach (with the exception of the 199 households in the case of income) were used by SALDRU to conduct poverty and inequality analyses (Argent et al. & Finn et al. 2009). They are also used in this dissertation to derive poverty and inequality estimates, unless stated otherwise.

[36] The question was asked as "What was the total amount of income (after income tax) that this household received last month? Please note this includes all the household members' salaries and wages, grants, interest, rental income and income from agriculture earned by household members in the last month."

[37] The question was asked as "Please would you look at the show card and point out the most accurate earnings category for last month's household income?" The fifteen categories are as follows: 1: None, 2: R1-R200, 3: R201-R500, 4: R501-R1 000, 5: R1 001-R1 500, 6: R1 501-R2 500, 7: R2 501-R3 500, 8: R3 501-R4 500, 9: R4 501-R6 000, 10: R6 001-R8 000, 11: R8 001-R11 000, 12: R11 001-R16 000, 13: R16 001-R30 000, 14: R30 001-50 000, 15: R50 001 or more.

2.8 All Media Products Surveys (AMPSs)

AMPS, used mainly for market research, has been conducted either once or twice a year by SAARF since 1975. It collects information on the usage of a wide range of household goods and services. Only one respondent aged 16 or above from each household, not necessarily the household head, was asked to take part in the survey.

2.8.1 Sampling design and sample size

The sample size in each survey ranged between 12 000 and 24 000, and the sample was stratified first by province and then by the size of the community, with the communities being categorized into metropolitan areas (250 000 or more inhabitants), cities (100 000 to 250 000 inhabitants), large towns (40 000 to 100 000 inhabitants), small towns (8 000 to 40 000 inhabitants), rural areas and farming communities. Large and small towns were grouped together, while rural areas and farming communities were grouped together. All metropolitan areas and cities were sampled. For the remaining community types, communities were ranked according to size then randomly selected.

Within the selected communities, interview points were randomly selected, with two addresses being selected at each interview point. If there was more than one household at any address, they were surveyed separately. Within households, one respondent was chosen using the Politz methodology. For instance, within each cluster of households, one male and one female were interviewed, each from separate households. Interviewers must return to a selected address three more times if their first attempt to survey the people residing there was unsuccessful. Only then will substitution be allowed (Van der Berg, Louw and Du Toit 2008).

Only people aged 16 years were asked to take part in AMPS. Moreover, the unique household number variable was not available, so it was not possible to identify participants from the same households, and the number of members taking part in the survey from each household was not known. Furthermore, to ensure that AMPS is a reliable data source, the data were validated externally using subscriber data from M-Net and new electricity connections made during the past 12 months from Eskom, after the weights were applied. Validation occurs by checking whether the AMPS figures fell within the 95% confidence intervals generated by these data sources. In addition, the data were internally validated using historical sales data for consumer durables with long lifespans and low duplication rates, such as microwaves and computers.

2.8.2 The questionnaire

2.8.2.1 Questionnaire structure

The survey mainly asked questions on product usage and product purchase during a certain period (e.g., within the last seven days or within the last four weeks).

2.8.2.2 Labour market status questions

As far as the labour market information is concerned, there was only one question asking the respondent to declare his/her work status, and the respondent could choose from the following options: (1) Working full-time, (2) Working part-time, (3) Not working – housewife, (4) Not working – A student, (5) Not working – retired, and (6) Not working – unemployed. In addition, the employed (i.e., those working full-time or part-time) were asked to declare the occupation, industry, whether they were self-employed or employees, and whether they were the main breadwinners of the household or not. Hence, there was insufficient information to determine the respondent's labour market status in both narrow and broad terms.

2.8.2.3 Household income and expenditure questions

Household expenditure was not captured, while the income information was collected through showing respondents cue cards divided into 29 or more categories, as shown in Table 2.23. The income question was asked as: "Please give me the letter which best describes the total monthly household income of all these people before tax and other deductions. Please include all sources of income, i.e., salaries, pensions, income from investments, etc."

If a respondent refused to answer the question on household income, SAARF imputed household income on the basis of household expenditure implied by the product questionnaire. Hence, in all AMPSs, no households had unspecified income. In addition, the category "R0" or "Zero income" was not included as an income category in all AMPSs. Furthermore, AMPS has an advantage that there are more income categories, and the income range in each category is narrower, compared with the categories in the censuses (Tables 2.4, 2.7 and 2.8) and OHSs/LFSs/GHSs (Tables 2.18 and 2.20). Whether the greater the number of intervals and the narrower width of each interval would improve the reliability of AMPS poverty and inequality estimates and trends would be discussed in the forthcoming chapters.

Table 2.24: Proportion of households in each monthly household income category, AMPSs

		1993		1994	1995	1996		1997	1998	1999		2000	2001
1	R1-R99	2.2%	R1-R99	1.9%	2.4%	1.5%	R1-R99	1.1%	1.4%	0.8%	R1-R199	2.1%	2.6%
2	R100-R199	4.1%	R100-R199	3.3%	2.8%	2.5%	R100-R199	2.2%	1.5%	1.5%	R200-R299	2.8%	2.5%
3	R200-R299	8.5%	R200-R299	5.0%	5.3%	4.4%	R200-R299	3.6%	3.5%	3.4%	R300-R399	3.6%	3.3%
4	R300-R399	6.2%	R300-R399	8.4%	9.7%	5.8%	R300-R399	4.1%	3.6%	3.7%	R400-R499	3.4%	3.1%
5	R400-R499	4.4%	R400-R499	4.2%	3.9%	8.4%	R400-R499	10.1%	9.1%	3.9%	R500-R599	10.8%	12.0%
6	R500-R599	6.3%	R500-R599	5.4%	5.0%	4.3%	R500-R599	4.7%	5.0%	11.0%	R600-R699	3.8%	3.4%
7	R600-R699	5.0%	R600-R699	5.9%	4.5%	4.2%	R600-R699	3.7%	4.0%	4.0%	R700-R799	2.6%	2.6%
8	R700-R799	2.9%	R700-R799	3.1%	3.3%	2.0%	R700-R799	1.7%	2.2%	2.0%	R800-R899	2.9%	3.5%
9	R800-R899	6.3%	R800-R899	6.1%	5.4%	7.2%	R800-R899	6.3%	6.4%	4.5%	R900-R999	2.9%	2.9%
10	R900-R999	4.5%	R900-R999	4.7%	4.2%	4.1%	R900-R999	4.0%	4.0%	3.1%	R1 000-R1 099	6.9%	7.5%
11	R1 000-R1 099	4.1%	R1 000-R1 099	3.8%	3.6%	3.9%	R1 000-R1 099	3.6%	3.8%	5.3%	R1 100-R1 199	2.3%	2.3%
12	R1 100-R1 199	3.9%	R1 100-R1 199	3.8%	4.0%	4.0%	R1 100-R1 199	4.5%	3.6%	4.2%	R1 200-R1 399	5.0%	3.5%
13	R1 200-R1 399	4.3%	R1 200-R1 399	5.2%	5.3%	5.2%	R1 200-R1 399	4.6%	5.1%	4.5%	R1 400-R1 599	4.7%	3.7%
14	R1 400-R1 599	2.8%	R1 400-R1 599	2.8%	3.4%	3.3%	R1 400-R1 599	3.3%	3.6%	3.3%	R1 600-R1 999	4.0%	4.6%
15	R1 600-R1 999	4.7%	R1 600-R1 999	5.0%	5.7%	5.0%	R1 600-R1 999	5.6%	5.2%	4.8%	R2 000-R2 499	6.5%	5.0%
16	R2 000-R2 499	5.4%	R2 000-R2 499	5.4%	5.2%	6.2%	R2 000-R2 499	6.0%	5.9%	6.0%	R2 500-R2 999	4.3%	4.6%
17	R2 500-R2 999	3.6%	R2 500-R2 999	3.5%	3.1%	3.2%	R2 500-R2 999	3.3%	3.3%	3.4%	R3 000-R3 999	5.5%	6.3%
18	R3 000-R3 999	5.1%	R3 000-R3 999	5.4%	5.4%	5.6%	R3 000-R3 999	6.1%	6.1%	6.4%	R4 000-R4 999	5.3%	4.8%
19	R4 000-R4 999	4.0%	R4 000-R4 999	4.0%	3.9%	4.3%	R4 000-R4 999	4.2%	4.5%	4.4%	R5 000-R5 999	3.6%	3.8%
20	R5 000-R5 999	2.7%	R5 000-R5 999	2.9%	3.0%	2.6%	R5 000-R5 999	3.0%	2.6%	2.8%	R6 000-R6 999	2.6%	3.0%
21	R6 000-R6 999	2.4%	R6 000-R6 999	2.7%	2.8%	2.8%	R6 000-R6 999	3.2%	3.3%	3.5%	R7 000-R7 999	2.5%	2.1%
22	R7 000-R7 999	1.7%	R7 000-R7 999	1.8%	1.6%	1.9%	R7 000-R7 999	2.1%	2.1%	2.3%	R8 000-R8 999	2.0%	2.0%
23	R8 000-R8 999	1.1%	R8 000-R8 999	1.1%	1.2%	1.4%	R8 000-R8 999	1.4%	1.5%	1.5%	R9 000-R9 999	1.1%	1.6%
24	R9 000-R9 999	0.8%	R9 000-R9 999	1.1%	1.3%	1.4%	R9 000-R9 999	1.8%	1.7%	1.8%	R10 000-R10 999	2.5%	2.2%
25	R10 000-R10 999	1.0%	R10 000-R10 999	1.2%	1.3%	1.3%	R10 000-R10 999	1.7%	1.9%	2.0%	R11 000-R11 999	0.9%	1.0%
26	R11 000-R11 999	0.3%	R11 000-R11 999	0.3%	0.4%	0.4%	R11 000-R11 999	0.5%	0.5%	0.6%	R12 000-R13 999	1.3%	1.6%
27	R12 000-R12 999	0.5%	R12 000-R13 999	0.7%	0.8%	0.9%	R12 000-R13 999	1.3%	1.4%	1.5%	R14 000-R15 999	1.3%	1.2%
28	R13 000-R13 999	0.2%	R14 000-R15 999	0.4%	0.5%	0.7%	R14 000-R15 999	0.7%	1.0%	1.1%	R16 000-R17 999	0.7%	0.9%
29	R14 000+	1.3%	R16 000+	0.9%	1.1%	1.6%	R16 000-R17 999	0.3%	0.4%	0.5%	R18 000-R19 999	0.4%	0.6%
30		100.0%		100.0%	100.0%	100.0%	R18 000+	1.3%	2.0%	2.2%	R20 000+	1.9%	2.1%
								100.0%	100.0%	100.0%		100.0%	100.0%

Table 2.24: Continued

		2002	2003	2004	2005	2006		2007	2008		2009
1	R1-R199	2.2%	1.9%	1.6%	1.7%	1.5%	R1-R299	1.4%	0.9%	R1-R499	3.2%
2	R200-R299	1.9%	1.5%	1.3%	1.0%	0.9%	R300-R399	1.1%	1.0%	R500-R599	1.8%
3	R300-R399	2.2%	2.2%	2.1%	1.7%	1.7%	R400-R499	1.0%	1.0%	R600-R699	1.5%
4	R400-R499	2.4%	2.1%	2.0%	1.8%	1.6%	R500-R599	1.4%	1.5%	R700-R799	1.4%
5	R500-R599	7.0%	3.1%	2.6%	2.5%	2.4%	R600-R699	1.5%	1.5%	R800-R899	2.5%
6	R600-R699	7.6%	5.0%	2.7%	2.4%	2.2%	R700-R799	1.8%	1.1%	R900-R999	3.5%
7	R700-R799	2.6%	8.3%	10.4%	8.7%	7.6%	R800-R899	7.8%	5.7%	R1 000-R1 099	7.4%
8	R800-R899	3.4%	3.6%	3.7%	3.4%	4.6%	R900-R999	2.1%	3.3%	R1 100-R1 199	2.2%
9	R900-R999	2.5%	2.4%	2.7%	2.6%	2.4%	R1 000-R1 099	3.8%	3.1%	R1 200-R1 399	3.4%
10	R1 000-R1 099	5.3%	4.7%	4.6%	4.0%	4.0%	R1 100-R1 199	1.6%	1.3%	R1 400-R1 599	3.1%
11	R1 100-R1 199	2.5%	1.8%	1.6%	1.5%	1.8%	R1 200-R1 399	3.4%	2.7%	R1 600-R1 999	4.4%
12	R1 200-R1 399	4.6%	3.9%	3.3%	3.6%	3.7%	R1 400-R1 599	4.2%	3.2%	R2 000-R2 499	6.4%
13	R1 400-R1 599	4.0%	6.4%	7.5%	7.4%	6.8%	R1 600-R1 999	6.9%	5.6%	R2 500-R2 999	4.5%
14	R1 600-R1 999	4.5%	4.8%	4.6%	4.9%	5.2%	R2 000-R2 499	6.0%	6.8%	R3 000-R3 999	7.4%
15	R2 000-R2 499	6.0%	6.3%	6.5%	5.9%	6.3%	R2 500-R2 999	5.9%	5.5%	R4 000-R4 999	5.5%
16	R2 500-R2 999	5.2%	5.4%	5.7%	5.2%	5.3%	R3 000-R3 999	6.6%	7.5%	R5 000-R5 999	5.0%
17	R3 000-R3 999	6.9%	6.8%	7.0%	6.9%	6.8%	R4 000-R4 999	5.6%	6.1%	R6 000-R6 999	4.4%
18	R4 000-R4 999	5.4%	5.2%	5.3%	5.3%	5.3%	R5 000-R5 999	5.3%	5.5%	R7 000-R7 999	3.7%
19	R5 000-R5 999	4.0%	3.8%	3.7%	4.5%	4.5%	R6 000-R6 999	3.6%	4.2%	R8 000-R8 999	3.5%
20	R6 000-R6 999	3.2%	3.1%	3.0%	3.5%	3.7%	R7 000-R7 999	3.0%	3.5%	R9 000-R9 999	2.2%
21	R7 000-R7 999	2.3%	2.2%	2.3%	2.9%	2.8%	R8 000-R8 999	3.5%	3.7%	R10 000-R10 999	3.7%
22	R8 000-R8 999	2.3%	2.3%	2.3%	2.8%	2.9%	R9 000-R9 999	2.5%	2.5%	R11 000-R11 999	1.6%
23	R9 000-R9 999	1.7%	1.5%	1.4%	2.0%	1.9%	R10 000-R10 999	3.9%	4.1%	R12 000-R13 999	2.4%
24	R10 000-R10 999	2.1%	2.2%	2.3%	2.8%	2.9%	R11 000-R11 999	1.6%	1.7%	R14 000-R15 999	2.8%
25	R11 000-R11 999	1.4%	1.2%	1.2%	1.3%	1.2%	R12 000-R13 999	2.6%	2.8%	R16 000-R19 999	3.0%
26	R12 000-R13 999	1.6%	1.8%	1.8%	1.9%	2.1%	R14 000-R15 999	2.9%	3.1%	R20 000-R24 999	3.5%
27	R14 000-R15 999	1.4%	1.6%	1.8%	1.9%	2.1%	R16 000-R19 999	2.6%	3.4%	R25 000-R29 999	2.2%
28	R16 000-R19 999	1.4%	1.4%	1.4%	1.9%	1.9%	R20 000-R24 999	2.8%	3.3%	R30 000-R39 999	1.9%
29	R20 000-R24 999	1.2%	1.4%	1.4%	1.4%	1.7%	R25 000-R29 999	1.4%	1.8%	R40 000-R49 999	1.0%
30	R25 000-R29 999	0.6%	0.8%	0.7%	0.9%	0.9%	R30 000-R39 999	1.0%	1.5%	R50 000+	0.9%
31	R30 000-R39 999	0.3%	0.6%	0.7%	0.7%	0.8%	R40 000+	1.3%	1.8%		100.0%
32	R40 000+	0.5%	0.7%	0.6%	1.0%	1.1%		100.0%	100.0%		
		100.0%	100.0%	100.0%	100.0%	100.0%					

2.9 Chapter summary

This chapter examined at the sampling methodology and the capture of labour market status, income and expenditure information in each survey. The sampling methodology differed amongst the surveys. OHSs, LFSs and QLFSs were the surveys that aim primarily at capturing labour market information in great detail. Table 2.24 summarizes the derivation of household income and expenditure in each survey. Both the income and expenditure data are available in the IESs, NIDS and PSLSD, and the data were captured in exact amounts. The aggregate amounts were derived by adding the income (or expenditure) amounts from different sources. IES 2005/2006 and NIDS explicitly took imputed or implied rent into consideration when deriving the total income. In contrast, Census and AMPS did not capture expenditure data, but only captured income data in bands.

Some of the OHSs and LFSs captured information on household expenditure. The respondents were asked to declare the exact amounts in OHS 1996-1998, but were asked to declare the relevant expenditure category in OHS 1999 and in the LFSs. OHS 1999 captured information on both household income and expenditure. Moreover, all GHSs only captured household expenditure in bands, and the bands were exactly the same as those asked in the OHS 1999 and the LFSs in nominal terms, except GHS 2009. Furthermore, in the two censuses, CS 2007, AMPSs, OHSs, LFSs and GHSs, the respondents were only asked to declare the aggregate income or expenditure amount or the relevant category from all sources.

For the surveys that captured income or expenditure data in bands, the number and width of bands differ amongst the surveys. The income bands are relatively wider, and the proportions of households with zero or unspecified income are much higher in the two censuses and CS 2007. Finally, IES 2005/2006 was the only survey adopting the diary method, but it was still complemented by the recall method.

To conclude, the income and expenditure information was collected quite differently amongst the surveys under study, with regard to the choice of the method to collect the information (Recall method vs. Diary method); whether the respondents were asked to report information in actual amounts or intervals; whether the total household income or expenditure amount was reported as a single estimate or derived as the aggregate of amounts from various sources, the number of intervals and width of each interval; and the proportion of households with zero or unspecified income or expenditure.

Table 2.24: A summary of Derivation of household income and expenditure in each survey: A summary

Survey	Year	Question asked?	Recall or diary method?	Data captured in bands or actual amounts?	'One-shot' overall amount / category or aggregation of amounts from different sources?	Number of bands, if the data are captured in bands
Income						
Census	1996 2001 2007	Yes	Recall	Bands[#]	Overall	Between 12 and 14
IES	1995 2000 2005/2006	Yes	Recall	Actual amounts	Aggregation	N/A
OHS	1995 – 1999	Yes (1999 only)	Recall	Bands	Overall	8
LFS	2000 – 2007	No	N/A			
QLFS	2008 – 2009	No				
GHS	2002 – 2009	No				
PSLSD	1993	Yes	Recall	Actual amounts	Aggregation	N/A
NIDS	2008	Yes	Recall	Actual amounts[##]	Aggregation[##] Overall[###]	15[####]
AMPS	1993 – 2009	Yes	Recall	Bands	Overall	Between 29 and 32
Expenditure						
Census	1996 2001 2007	No	N/A			
IES	1995 2000 2005/2006	Yes	Recall in 1995 and 2000; recall and diary methods in 2005/2006	Actual amounts	Aggregation	N/A
OHS	1995 – 1999	Yes (In 4 surveys)	Recall	1996 – 1998: Actual amounts 1999: Bands	Overall	8 (1999)
LFS	2000 – 2007	Yes (In 4 surveys)	Recall	Bands	Overall	8
QLFS	2008 – 2009	No	N/A			
GHS	2002 – 2009	Yes	Recall	Bands	Overall	Between 8 and 10
NIDS	2008	Yes	Recall	Actual amounts	Aggregation	N/A
AMPS	1993 – 2009	No	N/A			

[#] It is possible to derive the amount, and households in the same income band could have different income amounts.

[##] In 7 103 (out of 7 302) households, total household income equals the sum of wage income, government grant income, other government income, investment income, remittances income, implied rent income and agricultural income (See Section 2.7.2.3).

[###] In 185 households, total household income equals the 'one-shot' amount on household income plus implied rent income (See Section 2.7.2.3).

[####] In 14 households, total household income equals the mid-point amount of the relevant household income band plus implied rent income (See Section 2.7.2.3).

In addition, the weighting techniques were not the same across the surveys. How all the abovementioned issues could affect the comparability and reliability of poverty and inequality estimates and trends will be discussed in detail in Chapter 3. With regard to the labour market aggregates, the aforementioned mentioned inconsistent weighting techniques as well as the labour market status derivation methodologies adopted in the OHSs, LFSs and QLFSs would affect the validity of the labour market trends. The latter issue was only briefly dealt with in Section 2.4.2.2 but will be discussed in greater detail in Chapter 3.

CHAPTER THREE: FACTORS INFLUENCING THE RELIABILITY AND COMPARABILITY OF LABOUR MARKET, POVERTY AND INEQUALITY ESTIMATES ACROSS SURVEYS

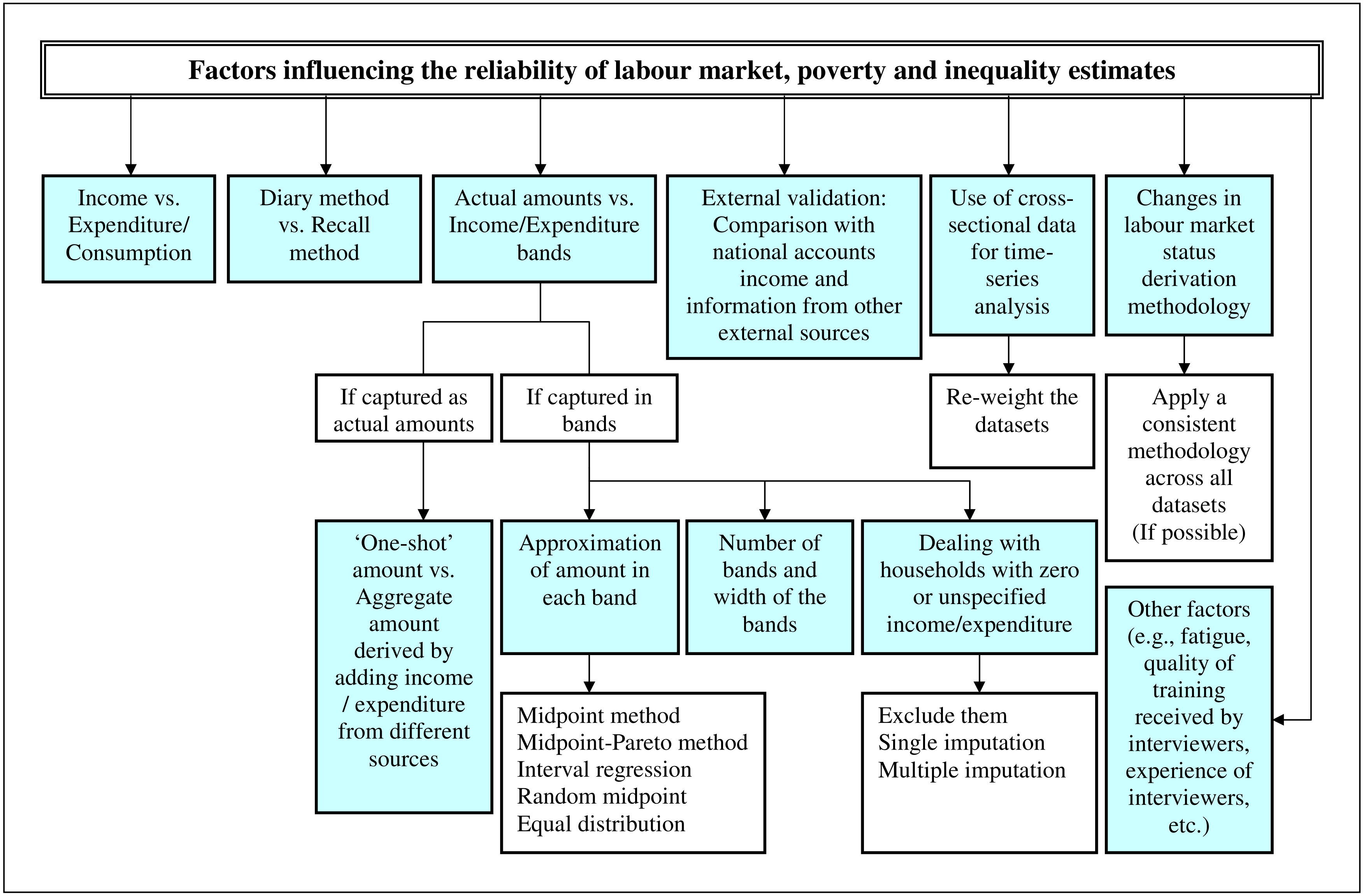
3.1 Introduction

One of the main reasons for collecting household survey data is to measure and understand the living standards of the population in a country. At the very least, such measurement requires data on money-metric variables such as income and expenditure (consumption). For broader concepts of living standards, information on non-money-metric indicators such as health, nutrition, life expectancy, educational attainment, literacy and housing is required (Deaton 1997: 22). With regard to the two money-metric variables, namely income and expenditure, as discussed in Chapter 2, the information was captured very differently in various surveys (See Figure 3.1), and this would lead to lots of comparability issues when one uses these two variables to examine poverty and inequality levels and trends. Hence, Chapter 3 will discuss how the methods used to collect income and expenditure information would influence the comparability and reliability of poverty and inequality estimates across surveys.

The accuracy of survey income and expenditure data could be assessed by comparing them with data from external sources like petrol consumption and income tax revenue of National Treasury. In addition, some argue that household surveys under-estimate income or expenditure, and hence the data should be adjusted (i.e., shifting the distribution rightwards) in line with the national accounts data. However, the other side of the argument is that national accounts data are also problematic. Therefore, these issues are examined.

With regard to the labour market information, Chapter 2 only briefly discussed that the methodology to derive both the narrow and broad labour market status has changed throughout the years, in all OHSs, LFSs and QLFSs under study. These changes will be discussed in depth in Chapter 3 to investigate how the comparability of labour market estimates could be influenced. The chapter concludes by discussing the cross entropy approach to re-weight the cross-sectional survey data to be consistent with demographic and geographic numbers presented by the ASSA and census data. The labour market, poverty and income distribution trends before and after the re-weighting will later be looked at in Chapters 4 and 6.

Figure 3.1: Factors influencing the reliability and comparability of poverty, inequality and labour market estimates, using survey data

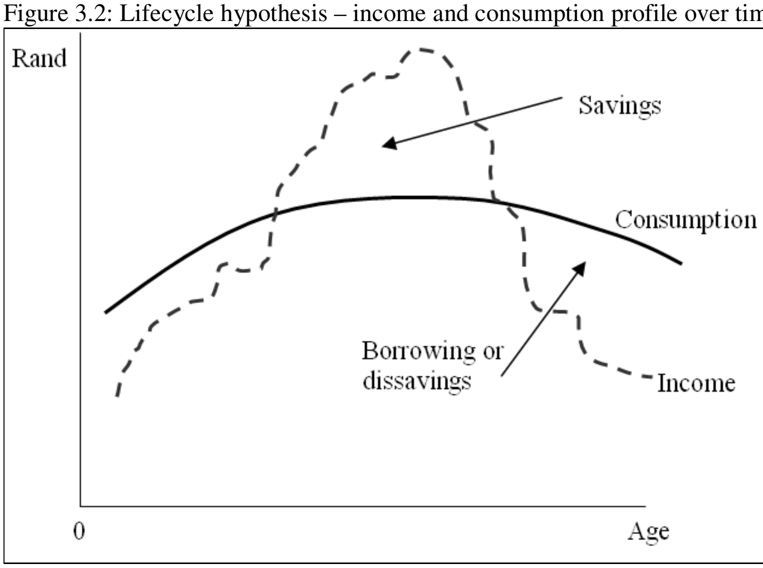


3.2 Income vs. Expenditure / Consumption

An important question that arises when using the money-metric approach to measure the standard of living, poverty and inequality of the population is whether income or expenditure / consumption should be used. The general argument (Haughton and Khandker 2009: 30) is that most rich countries use the income variable, as most of the income comes from salaries and wages and hence it is comparatively easy to measure, while it is difficult to quantify both the volumes and amounts of purchase when it comes to capturing expenditure. In contrast, in the poor countries, income is harder to measure as much of it comes from self-employment in informal activities, but expenditure is more straightforward and easier to estimate. Thus, consumption is the preferred variable. This section will analyse the advantages and disadvantages of using each variable in greater detail.

The primary reason to use the expenditure / consumption variable is that, in addition to fluctuating somewhat from year to year, income normally increases and then decreases in the course of a person's lifetime. In contrast, consumption remains relatively stable, since it could be smoothed to some extent by saving and borrowing (Blundell and Preston 1998: 603; McKay 2000: 85-86; Duclos and Araar 2006: 21; Haughton and Khandker 2009: 24-25). The permanent income hypothesis predicts this smoothing of short-term income fluctuations. Under this hypothesis, transitory (temporary) income is saved, while long-term (permanent) income is largely consumed, as shown in Figure 3.2.

Figure 3.2: Lifecycle hypothesis – income and consumption profile over time



Source: Haughton and Khandker (2009: 24).

Information on consumption over a relatively short period is more likely to represent a household's general level of welfare than the equivalent information on the more volatile income (Haughton and Khandker 2009: 25). Although random irregularities and seasonal patterns are present in consumption, they are normally smaller than those of income, as consumption is less tied to seasonal and weather-related patterns in agriculture than is income (Deaton and Grosh 2000: 93-94)

Secondly, the concept of consumption – giving money in exchange for a good or service – is clear to both interviewers and interviewees, while the income concept might not be clear (to be discussed later). Consumption is also more readily observed, recalled and measured than income (at least in developing countries, although this is not always the case) (Deaton and Grosh 2000: 93-94; Duclos and Araar: 2006: 21). Thus, it is easier to recall information on consumption. Finally, consumption is preferred over income as the former shows the current actual material standard of living by reflecting more directly the degree of commodity possession (Haughton and Khandker 2009: 30).

Using expenditure / consumption instead of income to measure money-metric poverty and inequality also has its drawbacks. First, there is a need for collecting data on consumption of goods and services item by item, in the case of aggregation approach. The number of consumption items could be as many as more than a thousand, while the income source items are much fewer. For example, in the 57-page IES 2000 questionnaire, only 6 pages were devoted to collecting information on income, while about 45 pages of the questionnaire were aimed at collecting consumption / expenditure information.

Although respondents are more likely to remember consumption activities in greater detail and to report higher expenditure if the questions are more detailed (Haughton and Khandker 2009: 25), such a longer questionnaire (e.g., if the aggregation approach is adopted) devoted to collecting consumption information is very costly and time-consuming. However, if a shorter questionnaire is used in order to save money and time (e.g., if the single estimate approach is adopted or if the respondents are only asked to report the overall food and non-food spending, instead of being asked to declare spending on each food and non-food item), this could result in inaccurate estimates of total consumption (Guenard and Mesple-Somps 2010: 523). In addition, the respondents might not provide answers to all consumption items or might not remember the amounts spent on all items, and so imputations have to be made (Deaton and Grosh 2000: 93-94). Furthermore, overly long recall periods (e.g., one year) could lead to under-estimation of consumption as memories fade as time goes by, i.e., recall bias arises (Guenard and Mesple-Somps 2010: 523), but longer recall periods might really be required for durable goods with low purchase frequency.

Households are also likely to under-report what they have spent on luxury or illicit items, e.g., alcohol, tobacco, or drugs. With regard to consumption on durable goods, such expenses are not regularly incurred, so the data could be noisy because recall bias is more likely to happen with longer recall periods. Looking at the durable goods consumption in greater detail, it is difficult to measure it, as it is not sure whether the full consumption amount on a durable good should be included, or whether only the change in the asset's value during the year (i.e., depreciation, plus the cost of locking up one's money in the asset) should be included[38] (Haughton and Khandker 2009: 25).

Deaton (1997: 32) argues that the presence of substantial inflation tends to overstate consumption relative to income, given that surveys usually have different reporting periods for the two variables. The reference period for consumption varies from item to item (e.g., the reference period could be one week, one month or one year for different items in IES 2000), while the importance of seasonality of income means that reference periods for income items are usually a year. Consumption is then denominated in more recent, higher prices than is income, resulting in an upward bias to measures of consumption. The following example could simplify the explanation: if a respondent takes part in the survey in December and is asked to declare the food consumption for the past month, and his answer is R100, then the annual food consumption is derived as R1 200 (R100 × 12 months). However, the same basket of food might be cheaper in the earlier months of the year, and if the respondent is asked to declare the annual food consumption (i.e., longer reference period) instead, the actual amount could be lower than R1 200 (providing the respondent remembers the food expenditure month by month clearly – keep in mind recall bias is more likely to happen with longer recall periods), i.e., the R1 200 amount derived using a shorter reference period might be biased upwards due to the impact of inflation.

Another disadvantage of using consumption relates to the difficulty of disentangling production and consumption (Deaton 1997: 28). As most agricultural households are both producers and consumers, they might find it difficult to distinguish consumption from production. In addition, home-produced items, typically food grown or raised on the farm or

[38] In all surveys under study, the full consumption amount on a durable good was included.

in kitchen gardens, should be properly recorded as both income and consumption, but are often very difficult to value.

Having discussed the pros and cons of using the consumption variable, there are also arguments for and against using the income variable. The main argument in favour of using income is that it is easier, cheaper and quicker to collect income data, especially in circumstances where income comes from only one or two sources (e.g., wages and pension) that are easily remembered or for which independent documentation exists (Deaton and Grosh 2000: 93-94). This is more likely to happen in richer, developed countries. Even if the household's income might come from many sources, it is still relatively easier to measure income than consumption, given the limited number of income sources (e.g., salaries and wages, pensions, remittances, interest received, income from businesses, etc.).

As far as the problems of using income are concerned, as mentioned previously (the lifecycle hypothesis), income of many households could be very volatile seasonally during the year, or from one year to another year, as a result of being subject to significant shocks. This is more likely to happen in households engaging predominantly in self-employment, or agricultural activities or households that are heavily reliant on transfers from either public or private sources. As a result, measuring the household annual income might require a lot of visits to the household or dependence on the ability of households to remember their income from many months earlier (Deaton and Grosh 2000: 93-94; McKay 2000: 84-86). In addition, as a result of the volatile nature of income (the income could be temporarily high or low), the reporting period might not be able to capture the 'mean' income of the household accurately.

The concept of income, especially income from self-employment or own-account agricultural and informal activities, is often unclear (Deaton and Grosh 2000: 93-94). Respondents might not genuinely know how much income they make in these activities, in particular due to reasons like seasonal variations, income declarations are biased by under-declarations and non-responses, or respondents might not feel there is a need to report incomes earned infrequently or might not consider them as part of income, e.g., receipt of transfers and remittances and other non-labour income (McKay 2000: 95; Haughton and Khandker 2009: 30; Guenard and Mesple-Somps 2010: 527). In addition, although income sources / items are fewer compared with consumption sources / items as mentioned above, Riphahn and Serfling (2004) argue that income sources are diverse, especially among the professionally self-employed in rich countries. A greater cognitive requirement on the respondent to provide information could result in lower response rates or the reporting of unreliable income information.

Posel and Casale (2005: 4) argue that each member of the household is more knowledgeable about his/her own income than about the income of the other members. Hence, item non-response – not specifying the household income – is significantly higher for proxy-reporting (i.e., only the household head or one member of the household is asked to declare the total household income earned by all members of the household, as in the OHSs, LFSs, GHSs and AMPSs) than for self-reporting (i.e., each member of the household is asked to declare his/her personal income, before the personal incomes of all members are added to derive household income, as in the two censuses and CS 2007)[39].

It is also argued by various researchers (Deaton 1997; Deaton and Grosh 2000; McKay 2000; Posel and Casale 2005; Haughton and Khandker 2009) that respondents are more likely to report inaccurate information about their income or refuse to declare the full extent of their income, as income is a more sensitive topic to ask about than consumption. This could be due to the fact that, as income is taxable in almost all countries, it is difficult for interviewers to convince respondents that the information they provide will not be passed on to the tax authorities. As a result, income would be reported inaccurately or understated.

Some respondents might be hesitant to report income earned illegally, such as smuggling, corruption or prostitution, and income earned from informal activities not reported to the tax authorities, such as street vending. Another reason the respondents might feel sensitive to disclose income information is that, income from assets is harder to capture, with the wealthy being typically thought to be less likely to co-operate as they might fear governmental or other uses of the data. In contrast, low-income earners might overstate their income, as they might think that by reporting low earned income they are considered being unsuccessful.

Chapters 5 and 6 will compare the income and expenditure data across the surveys to investigate if expenditure was captured better than income (with South Africa being a developing country) and if the poverty and inequality estimates are influenced as a result.

[39] However, this is not the case in the South African surveys. As discussed in Chapter 2, a fairly high proportion of people did not specify their personal income in the two censuses and CS 2007 (See Tables 2.3 and 2.6). This consequently resulted in a higher proportion of households with unspecified household income (See Tables 2.4, 2.7 and 2.8). In OHSs, LFSs and GHSs, only the household heads were asked to report household income or expenditure, and the proportion of households with unspecified answers (See Tables 2.18 and 2.20) was lower than in the censuses and CS 2007.

3.3 Diary method vs. Recall method

Regardless of whether income or expenditure (consumption) is chosen to measure poverty and inequality estimates, an important issue is how to collect the information. In all South African surveys under study, the recall method was adopted, except in IES 2005/2006, which adopted the diary method for the first time. The recall method is problematic for various reasons. First, recall bias is very likely to happen, as the respondents could not remember many purchases long after they have been made. This is likely to result in either an under-estimation of consumption (as it is not easy for people to remember their consumption from long ago) or to inaccurate guesses (i.e., respondents estimate their consumption over the whole year from their current rate of consumption) (Deaton 1997: 24-25; Deaton and Grosh 2000: 109-110). This recall bias becomes more serious as the recall period increases.

The telescoping phenomenon – respondents tend to include consumption events that took place before the beginning of the recall period (Deaton and Grosh 2000: 110) – is also likely to happen under the recall method. As a result, consumption could be over-estimated. For instance, when asked about expenditures during the previous year, respondents might include items they bought 13 months ago. Deaton and Grosh argued further that telescoping is more likely to happen in durable goods purchases and/or if the recall period becomes longer, since respondents are more likely to forget the date the consumption events occurred. Hence, for example, if a household taking part in the survey in October 2009 purchased a personal computer worth R5 000 in September 2008 (i.e., more than a year ago), but wrongly thought that it was bought in October 2008 and included it as part of expenditure for the recall period, this would result in the over-estimation of total expenditure.

Deaton (2005: 16) suggests a shorter recall period for accuracy of memory. Moreover, if the respondents' memories of their consumption fade quickly, many visits might be required throughout the year to ensure that data on high-frequency non-durable purchases are collected accurately, but the resultant increase of the frequency of the survey could be costly. In contrast, as the consumption of some items might only take place occasionally during a year, a longer recall period is required, e.g., consumption of durable items like motor vehicles might not have taken place in the last month but rather in the last year.

Looking at the issue of recall period further, the match between consumption and purchases is more accurate when averaged over a longer recall period (Deaton 2005: 16). For example, if the respondent is asked to declare consumption on food in the past month and the respondent takes part in the survey in December, it is likely that his/her food expenditure is higher than usual due to the festive season, and the resultant annual food expenditure derived from this monthly expenditure could be over-estimated. However, if the respondent is asked to declare the total food expenditure in the past 12 months, the seasonal fluctuations (i.e., food expenditure is lower at the start of the year but then higher in certain months) might be considered by the respondents (providing he/she remembers the monthly food expenditure with good memory), and the resultant food expenditure could be more accurate.

As a result of the drawbacks of using the recall method as discussed above, the diary method becomes an alternative approach to collect income and consumption information. Corti (1993) argues that it is a reliable alternative to the conventional interview method (which adopts the recall approach) for events that are easily forgotten or difficult to recall correctly; the diary method is designed to minimize dependence on respondents' memories and consequently reduces the likelihood of recall bias, especially on frequently purchased (non-durable) items which are normally more difficult to recall, since consumption events are recorded as they take place or close to that time (Deaton and Grosh 2000: 109; Battinstin 2003: 2; Wiseman et al. 2005: 395). The diary method is also more convenient to the respondents, as they could answer the questions at a time and place that are suitable for them (Deaton and Grosh 2000: 119-122; Wiseman et al 2005: 395).

The diary method also helps to reduce the problems associated with gathering sensitive information by personal interviews. For example, the respondent might feel uncomfortable if he/she is asked by the interviewer to recall total consumption on items like alcohol and tobacco, but will feel more comfortable to report the consumption on these items on a diary without the presence of the interviewer. Finally, diaries allow for the analysis of events over time (Wiseman et al. 2005: 395). For instance, it is possible to look at the effect seasonality has on expenditure, particularly in poor rural communities, if the diary method is adopted[40].

Despite its merits discussed above, the diary method is associated with various problems. First, diaries are less suitable where literacy levels are low, because the diary keepers might not be able to write down the purchase items correctly if given an unstructured diary so as to enter consumption activities on a blank page (this is the case in the IES 2005/2006 diary

[40] It is also possible to observe this seasonality effect in the recall period, providing the respondents are, for example, asked to declare expenditure on the items in each of the last 12 months. However, this approach was not adopted in all surveys under study.

approach, as the respondents were asked to describe the items, place of purchase and the consumption value on the weekly diary). Even if the diary is structured like a questionnaire in which the participants are only required to tick the printed boxes containing the consumption events and fill in the consumption amounts, some of them might not be literate enough to understand the meaning of these consumption items (Wiseman et al. 2005: 396). Hence, the data collected from the diaries might be biased towards the competent, literate diary keepers (Corti 1993). A pictorial diary might be required eventually to improve the accuracy of the responses of the people with lower literacy levels.

Deaton (2005: 16) and Wiseman et al. (2005: 399-400) argue that the diary method might not suit the more diverse, well-off households with bigger household size; if the responsibility for spending lies with more than one person in the household, individuals have insufficient knowledge of what each household member spends. Moreover, some family members are outside home most of the time, multiple diaries per household should be considered, but it would become much more costly and time-consuming to collect and edit the information. Consequently, overlap in entries made by different family members could happen.

If the households are asked to keep the diaries for a very short period of time (e.g., one week, or four weeks in the case of IES 2005/2006), the resultant consumption estimate might be inaccurate, as some households have unusually low purchase rates in certain items (e.g., every month or every few months, especially the semi-durable and durable goods, as discussed in Section 3.2). Hence, the diary method might work better for non-durable items as the purchases of these items take place more frequently; the recall method might work well to record the consumption of the more durable, bulky items with low purchase frequency (Deaton and Grosh 2000: 119-122; Battinstin 2003: 2), despite the fact that recall bias is more likely to happen in the latter approach due to the longer reference period required. This might explain why the recall method (questionnaire) was still used in IES 2005/2006 to complement the diary method, with the recall method focusing on collecting information on income as well as semi-durable and durable goods consumption[41], and the diary method primarily concentrating on the collection of non-durable consumption information (Table 2.14).

[41] This implies that inaccuracy in the durable goods consumption data is inevitable to a certain extent, regardless of which method is adopted: if the recall method is adopted, a longer reference period is required to collect reliable information since such consumption happens only occasionally, but a longer reference period is associated with a greater likelihood of recall bias and telescoping. If the diary method is adopted, durable goods consumption might be reported as low as zero. It is because the participants are only asked to keep the diaries for a few weeks and durable goods consumption might not have taken place at all during the diary-keeping period. However, when comparing the two approaches, it seems the recall method is the relatively better approach to collect information on durable goods consumption.

Telescoping and recall bias, as discussed previously, could still happen even if the diary method is adopted, despite the fact that the likelihood of it happening becomes lower, as the diaries still rely on the respondents' memory and might not be filled out every day (Deaton and Grosh 2000: 119-122). The chance that these two problems would occur increases if entries are not made as close as possible to the time of actual expenditure, since the respondents are left to their own devices to complete the diary and there is no guarantee that the respondents would report events immediately after they took place (Deaton 1997: 24-25 & Wiseman et al. 2005: 398). For example, if the respondent purchased various goods at a supermarket one day but the entries were only made on the dairy a few days later, consumption amounts might not be recalled correctly and the consumption of some goods might be forgotten and eventually not entered at all on the diary. Hence, the researchers might need to visit the households frequently to actively encourage them to regularly update the diaries. If it is found that there are missing data (e.g., consumption items are entered on the diary but the amounts spent are not reported), then the researchers have to go back to the respondents to clarify entries, but the data would eventually become retrospective and once again subject to recall bias (Wiseman et al. 2005: 395).

Corti (1993), Deaton and Grosh (2000: 119-122), Wiseman et al. (2005: 395) and Ahmed et al. (2006: 9-10) argue that the 'first-day effect' is likely to happen in the diary approach: the first day and first week of diary keeping shows higher reporting of consumption than the following days/weeks. It could be explained by various factors: the novelty of diary keeping wears off as time goes by; the respondents feel exhausted to keep records and eventually become less detailed in their reporting; the diary keepers no longer carry their diaries with them[42]. This is why, as mentioned above, intermediate visits from the interviewers or even incentives are required to preserve good diary keeping until the end of the period.

The recording of the use of illicit drugs or income that has not been declared to tax authorities might remain inaccurate under the diary method, even though it does not involve face-to-face communication as it happens in the recall (interview) method, as the respondents could still feel sensitive to enter such information on the diaries, and eventually decide not to fill in the above information at all on the diaries (Wiseman et al. 2005: 395).

[42] This could explain why only households that completed at least two weekly diaries were accepted in IES 2005/2006 (Table 2.13).

Although the diary method reduces the duration that the interviewer spends interviewing the households, this method might increase the time that the interviewer must spend travelling, as it requires additional trips to collect the completed diary. Moreover, considerable time might also be spent assisting illiterate households fill out the diaries. Furthermore, the interviewers might also need to visit the households frequently to examine the diary briefly, or to prompt the respondents to fill it out more completely if the diary appears to be incomplete. Consequently, the diary method could become more time-consuming to the interviewers compared with the recall method, might transform the situation back into an interview, and could even affect the motivation and competence of the interviewers due to reasons like fatigue (Corti 1993; Deaton and Grosh 2000: 119-122).

The diary method could be time-consuming and expensive (Sudman and Ferber 1971: 726; Corti 1993; Wiseman et al. 2005: 395): time is required to train the diary keepers and to maintain their support; intensive labour work is required to collect, edit and analyse the sheer volumes of data, especially if the diary is unstructured, since intensive editing and coding will push up the costs and involve even more time; respondents might be more co-operating and fill in the diaries more accurately only if offered incentives or gifts.

As discussed in Chapter 2, IES 2005/2006 is the only survey that adopted the diary method (to complement the recall method) to capture expenditure information (refer to Section 2.3, in particular Table 2.14). Chapter 5 will return to this issue by examining whether the expenditure data captured in this survey is particularly different when compared with IES 1995 and IES 2000 (which adopted the recall method only), whether the expenditure information is comparable across the three surveys, and most importantly, how the reliability and comparability of poverty and inequality estimates are affected as a result of the use of the diary method in IES 2005/2006.

3.4 Actual amount vs. Bands

As discussed in Chapter 2, participants in the surveys were asked to declare the exact income and expenditure (consumption) amounts in some surveys, or the relevant income and expenditure category in other surveys. An important question that arises is which method is more appropriate to collect the information better. Posel and Casale (2005: 10), Von Fintel (2006: 1) and Malherbe (2007: 25) argue that two major reasons the respondents did not declare the exact amounts in the surveys are that they are reluctant to disclose such information due to confidentiality or privacy concerns, and that they really do not know exactly how much they or other members in the households earn and/or spend. As a result, this leads to a high proportion of households with unspecified income or consumption information and also possible bias in the data collected.

Hence, respondents, especially those in the higher income / consumption categories, might prefer the anonymity of indicating to what predefined income / consumption interval (band) they belong. In fact, Posel and Casale (2005) found that, with regard to the information on income from the main job in LFS 2002b, bracket values instead of the actual amounts were more likely to be reported among those employed who are older, more educated, white, residing in urban areas, self-employed, informally employed and staying in larger households[43]. Von Fintel (2006) also found that people with higher earnings from the main job in LFS 2003b were more likely to report the relevant income category. Hence, it is suggested that the 'income bracket option' question be asked along with the 'exact income amount' question in order to boost the response rate and obtain more reliable income or expenditure information. This is not the case in all surveys under study, except the income information in NIDS (See Section 2.7.2.3).

Furthermore, this income band approach permits respondents to report with a margin of error, especially if they do not know the exact amounts earned. For example, if someone taking part in Census 1996 did not remember clearly his/her nominal personal income was R4 450.75, but he/she still remembered his/her income was somewhere between R4 400 and R4 500, then he/she would report his income to be under the "8: R3 501 – R4 500" interval. If he/she was only asked to declare the exact amount, he might end up refusing to answer this question, which would eventually cause his/her household income to be unspecified (See three decision rules to derive household income in Section 2.2.2.3). As a result, a significant greater response for income variables could be achieved and a better dataset with possibly more correct results created, if the interval approach is adopted.

A final problem of using the interval approach is that, as survey years progress, income brackets will invariably change with inflation. Alternatively, if the brackets are left unadjusted, an increasing proportion of households would fall in the higher-income categories due to the impact of inflation. Section 3.6 will return to this interval approach issue.

[43] Note that with regard to the question on income from the main job in the LFSs, the respondents were given two options to declare the income – either the exact amount or the relevant income category

3.5 Actual amount: Single-estimate amount or aggregation of amounts from sub-items

If income and expenditure information is to be collected by asking the respondents to declare the exact amounts earned or spent, the next issue to decide is whether to ask the respondents to declare the 'one-shot', single estimate (by asking questions like "What is the total income you earned from all sources in the past 12 months?" and "How much do you spent on all items in the past month?") as in OHS 1996-1998 and NIDS, or to aggregate the amounts from sub-items (i.e., by asking questions like "How much do you earn from income source X?", "How much do you earn from income source Y?", and so forth, and then the total income is derived by adding the amounts from the answers of these questions, as in the three IESs and NIDS.

The 'one-shot' amount, single estimate approach, despite being a relatively less time-consuming and costly method to collect the required information, could confuse the respondents, as they are unsure about what items should be included as part of the total income or expenditure. This may result in low response rate, and/or under-reporting of total income or expenditure (Deaton 1997: 27; Browning et al. 2002: 7-10). Hence, there is a need to disaggregate to some extent so as to obtain more satisfactory estimates.

If a series of questions are asked on all of the sub-items in order to derive the overall income or expenditure amount, an issue to consider is the appropriate level of disaggregation. Deaton (2005: 16) claims that the greater the degree of disaggregation of the number of items that are separately distinguished, the more accurate is the measured consumption (expenditure) in total. However, Deaton (2005: 16) as well as Browning et al. (2002: 12-18) suggest that, if the level of disaggregation is too high (with IES being a South African example), it could be very demanding, time-consuming and exhausting to both the interviewers and interviewees, and the latter might end up deliberately providing misleading amounts and even not answering some questions (i.e., item non-response). This eventually results in the derivation of an even more inaccurate aggregate consumption amount, compared with the single estimate method.

In addition, Browning et al. (2002: 19) argue that, for non-durable items, a non-exhaustive list method should be more than enough to obtain reliable information on consumption (expenditure), e.g., the two questions "expenditure on food at home" and "expenditure on food outside home" should result in a pretty good predictor of total food expenditure. In contrast, for durable items, they suggest that the exhaustive method works better.

With regard to the derivation of the aggregate income, Davern et al. (2005: 1535) claim that the 'one-shot' amount approach might work better, as asking respondents to declare exact amounts earned from each income source could be burdensome and intrusive, because people generally do not like to divulge how much money they earn in too great detail, as a result of the questions' sensitive nature. In fact, some respondents already find it disturbing to reveal income or even consumption information even if asked to declare the 'one-shot' amount.

As mentioned in Section 2.7, NIDS is the only survey that asked the respondents to report the household income and expenditure in actual amounts under both single estimate and aggregation approaches. Chapters 5 and 6 will return to this comparability issue to examine whether the more precise, aggregation approach leads to a higher income or expenditure amount (or whether the single estimate approach under-captured income or expenditure) and subsequently lower poverty levels and more accurate inequality estimates.

3.6 Allocation of amount in each band

If the income or expenditure information was collected in bands, the data needs to be made continuous before dividing it by household size to derive the per capita income or expenditure variable required for poverty and inequality analyses. Hence, the income or expenditure amount of each band needs to be determined. This section discusses the commonly used approaches to deal with this problem.

3.6.1 Midpoint method

The midpoint method is simple and widely used. In this method, each household who supplies its income / expenditure bracket is assumed to earn / spend the category mean – its midpoint. For example, if a household taking part in AMPS 2009 declares its nominal monthly household income falls in the "R5 000 – R5 999" category, the income amount is derived as R5 500. Similarly, if a household participating in GHS 2009 claims its nominal monthly household expenditure falls in the "R5 000 – R9 999" category, then the expenditure amount is approximated as R7 500. As far as the top category is concerned, since no upper limit exists, it is often assumed that the mean exceeds the lower limit by 10% (Fields 1989). For instance, if the nominal monthly household income category of a household from the AMPS 2000 sample is "R20 000+", the income amount is equal to R22 000 (R20 000 × 1.1).

Although this method lacks theoretical backing (Whiteford and McGrath 1994: 28), it may be attractive because of its simplicity. However, Seiver (1979: 230) is concerned that the true mean of any interval will always be below its midpoint, regardless of the number and width of the intervals, given intervals starting with "0"[44], as reported earnings or incomes tend to heap at levels ending in "0", or to a lesser extent, "5". For example, if the Census 1996 income categories were given as "R1 000 – R1 499", "R1 500 – R2 499" and so forth, then the people earning R1 500 would fall in the latter category, while the former category would be dominated by people earning R1 000. As a result, the true mean of the "R1 000 – R1 499" category would be smaller than its midpoint (R1 250). However, if the categories were given as "R1 001 – R1 500", "R1 501 – R2 500", etc., (i.e., ending in "0") like they were asked in Census 1996, the former category would probably be dominated by people stating they earned R1 500, and the true mean of this interval would exceed the midpoint (R1 250).

3.6.2 Midpoint-Pareto method

As the lower income / expenditure categories are narrow (as is the case in the surveys under study, as discussed in Chapter 2), Whiteford and McGrath (1994: 29) argue that the distribution of income at the bottom end is not noticeably influenced by midpoint imputation. However, as greater skewness within groups becomes evident in the higher income categories, a parametric approach is necessary there. A Pareto mean can be estimated for the open interval. This value could deviate from the midpoint, according to the heaviness of the tail (Von Fintel 2006: 15).

The Pareto mean (in the case of household income) is calculated as follows (Cloutier, 1988: 417; Gustavsson 2004: 20; Whiteford and McGrath 1994: 83):

- A Pareto function is fitted to the data by regressing $\log N$ against $\log Y$, i.e., $\log N = c + \alpha \log Y$, where Y stands for the lower limit of a household income interval and N represents the number of households with the household income above Y;
- Successive regressions are conducted each time eliminating the lowest income interval, until the highest coefficient of determination (R^2) is found, subject to the constraint that no less than the last three intervals before the open interval are used;
- The Pareto coefficient (α) from the chosen Pareto function is used in this equation to calculate the means of each of the bounded income intervals: $\bar{x} = \left[\frac{\alpha}{\alpha+1}\right] \cdot \left[\frac{x_1^{\alpha+1} - x_2^{\alpha+1}}{x_1^{\alpha} - x_2^{\alpha}}\right]$,

[44] This is the case in the OHSs/LFSs, GHSs and AMPSs.

where x_1 and x_2 are the upper and lower bounds of the interval;

- The Pareto coefficient is also used to calculate the mean of the open interval. That is, $\bar{x} = \left[\frac{\alpha}{\alpha+1}\right] \cdot x_\infty$, where x_∞ represents the lower limit of the open interval.

The midpoint-Pareto method is applied in the categorical data in either of the following ways: (1) the midpoint is used for all categories except the open category, while the Pareto method is applied to derive the Pareto mean for the latter category; (2) the midpoint is used for categories up to and including the category containing the population median income, and the Pareto mean is used for categories above the aforementioned category.

Table 3.1: Applications of midpoint and midpoint-Pareto methods on Census 1991

	Mean of each interval		
Nominal monthly household income	**Midpoint method**	**Midpoint-Pareto method (1)** #	**Midpoint-Pareto method (2)** ##
1: No income	R0	R0	R0
2: R1 – R499	R250	R250	R250
3: R500 – R699	R600	R600	R600
4: R700 – R999	R850	R850	R850
5: R1 000 – R1 499	R1 250	R1 250	R1 250
6: R1 500 – R1 999	R1 750	R1 750	R1 750
7: R2 000 – R2 999	R2 500	R2 500	R2 500
8: R3 000 – R4 999	R4 000	R4 000	R4 000
9: R5 000 – R6 999	R6 000	R6 000	R6 000
10: R7 000 – R9 999	R8 500	R8 500	R8 500
11: R10 000 – R14 999	R12 500	R12 500	R12 500
12: R15 000 – R19 999	R17 500	R17 500	R17 106
13: R20 000 – R29 999	R25 000	R25 000	R23 899
14: R30 000 – R49 999	R40 000	R40 000	R37 253
15: R50 000 – R69 999	R60 000	R60 000	R58 163
16: R70 000 – R99 999	R85 000	R85 000	R82 083
17: R100 000 – R149 999	R125 000	R125 000	R119 495
18: R150 000 – R199 999	R175 000	R175 000	R171 061
19: R200 000 – R299 999	R250 000	R250 000	R238 990###
20: R300 000 – R499 999	R400 000	R400 000	R372 531###
21: R500 000 or above	R550 000	R880 193	R880 193###

Source: Whiteford and McGrath (1994: 84).

Method 1: Midpoint is used for all categories except the open category, while the Pareto method is applied to derive the Pareto mean for the latter category.

Method 2: Midpoint is used for categories up to and including the category containing the population median income, and the Pareto mean is used for categories above this category.

The mean of the R200 000 – R299 999 interval = $\left[\frac{-2.3151}{-1.3151}\right] \cdot \left[\frac{299999^{-1.3151} - 200000^{-1.3151}}{299999^{-2.3151} - 200000^{-2.3151}}\right] = $ R238 990,

while the mean of the R300 000 – R499 999 interval = $\left[\frac{-2.3151}{-1.3151}\right] \cdot \left[\frac{499999^{-1.3151} - 300000^{-1.3151}}{499999^{-2.3151} - 300000^{-2.3151}}\right] = $ R372 531.

The Pareto mean of the open interval is derived as: $\left[\frac{-2.3151}{-1.3151}\right] \cdot 500000 = $ R880 193.

Whiteford and McGrath applied midpoint-method (2) on the Census 1991 household income data. The Pareto equation from the regression of 10 observations was: $\log N = 14.04 - 1.938 \log Y$, where -1.938[45] was the Pareto coefficient. Table 3.1 above presents their results and the means of each interval had midpoint-Pareo method (1) been applied[46].

In the South African studies (to be discussed in Chapter 5), midpoint-Pareto method (1) is the commonly used approach. For the remainder of the dissertation, this method will be used to derive the midpoint income or expenditure amount of the interval data, and will be simply referred to midpoint-Pareto method.

3.6.3 Interval regression

Interval regression tries to predict the income (or expenditure) amount from some well chosen explanatory variables, such as educational attainment, age, gender, race, labour market status of household head, household size, number of employed members in the household, etc. The lower limit and upper limit of each income or expenditure category (with the exception of the open interval – there is no upper limit) must be specified in the interval regression, before the model could predict what income / expenditure each household will earn / spend based on the explanatory variables used.

3.6.4 Random midpoint method

This method uses the midpoint of an income / expenditure interval and then distributes the households falling within the income / expenditure level randomly across the interval. If f_i stands for the frequency of households falling within income level i and x_i represents the midpoint of income level i, the following model is applied to obtain the random midpoint dataset (Malherbe 2007: 37): $Y_{ij} = x_i + sign_{ij} \times U_{ij}(0, x_i - lower\ limit_i)$ where Y_{ij} is the new random midpoint income value for income level i and household j, $j = 1,2,..., f_i$, $sign_{ij}$ is the sign for income level i and household j, where $sign_{ij}$ has a 50% chance of being +1 and 50%

[45] However, it is suspected that the authors did not use this coefficient to derive the Pareto mean. In fact, the coefficient should be -2.3151 (instead of -1.938).

[46] From the last column of Table 3.1, the Pareto mean was derived from the category '11: R10 000 – R14 999' onwards. However, it is unlikely for the 1991 median monthly household income to fall in this range. For instance, the median monthly household income in Census 1996, Census 2001 and CS 2007 are about R16 000, R13 000 and R19 000 respectively (in 2000 prices). Using these three amounts, the median income in 2000 prices in Census 1991 is ranged between R6 364 and R9 300, i.e., falling in either the '9: R5 000 – R6 999' or '10: R7 000 – R9 999' categories. Thus, it is not sure if Whiteford and McGrath (1994) derived the Pareto mean from the interval containing the median income or rather from the median income interval onwards.

chance of being –1, and U_{ij} is the uniform distribution, with lower limit of 0 and upper limit of $x_i - lower\ limit_i$, where *lower limit*$_i$ represents the lower limit of income level i.

For example, if a household fell in the "R400 – R799" monthly household expenditure category in GHS 2008, the midpoint (i.e., x_2) is R600, while the lower limit of this interval (i.e., *lower limit*$_2$) is R400. Assuming $sign_{ij}$ is –1 for this household, and a random draw from the uniform distribution (lower limit and upper limit being 0 and 200 respectively) gives an amount of R50, then the estimated household expenditure amount is derived as: $600 + (-1) \times 50 = \text{R550}$. Similarly, using the same information but if $sign_{ij}$ is +1 for this household, the household expenditure is calculated as: $600 + (+1) \times 50 = \text{R650}$.

3.6.5 Equal distribution method

This method assumes that income recipients are equally distributed within each category. For example, if 400 households fell in the "2: R400 – R799" monthly household expenditure category in GHS 2008, the first randomly chosen household from this interval is assumed to have monthly expenditure of R400, the second and third randomly chosen households are assumed to have monthly expenditure of R401 and R402, and so forth, and the 400-th and the last randomly chosen household is supposed to spend R799. However, the method is cumbersome since it generates a huge number of records (Whiteford and McGrath 1994: 30), as the width of the interval and the number of households falling in the interval increase.

Having discussed the various methods to derive the income / expenditure mean of each interval, the comparability of the results of these methods, as well as the quality of the data captured in the actual amount method and the interval method are considered. In South Africa, Von Fintel (2007), who looked at the LFS 2003b data on earnings from the main job and applied various methods (midpoint method, mid-point Pareto method, interval regression and lognormal distribution) to make the categorical earnings data continuous, found that coefficients of the Mincerian earnings regressions were invariant to the methods used. His study did not investigate the impact of each method on poverty and inequality estimates. In contrast, Malherbe (2007) applied the Census 2001 income intervals to the IES 2000 data, and separately applied the midpoint method, interval regressions method and random midpoint method to derive the amount in each category. At the end, Malherbe found that the poverty and inequality estimates were very similar for the continuous and midpoint data, while the interval regressions and random midpoint method obtained different results. The interval regression data under-estimated poverty, while the results obtained from the random midpoint data were not usable and the data were eventually rejected by Malherbe.

3.7 Number of bands and width of each band

If the respondents in a survey report their income or expenditure by declaring the relevant category, one might be concerned that the results of the poverty and inequality estimates would be heavily influenced by the number and width of the income / expenditure bands of the survey concerned. From Table 3.2, it could be seen the number of bands is as few as eight in the OHSs/LFSs/GHSs but as many as 32 in the AMPSs. The width of the bands ranges from R100 (e.g., in AMPS 2009) to R102 400 (e.g., in Census 2001 and CS 2007).

Table 3.2: Number and width of income and expenditure bands in selected surveys

Census 1996 – Income	Width	AMPS 2009 – Income	Width
R1 – R200	200	R1 – R499	500
R201 – R500	300	R500 – R599	100
R501 – R1 000	500	R600 – R699	100
R1 001 – R1 500	500	R700 – R799	100
R1 501 – R2 500	1 000	R800 – R899	100
R2 501 – R3 500	1 000	R900 – R999	100
R3 501 – R4 500	1 000	R1 000 – R1 099	100
R4 501 – R6 000	1 500	R1 100 – R1 199	100
R6 001 – R8 000	2 000	R1 200 – R1 399	200
R8 001 – R11 000	3 000	R1 400 – R1 599	200
R11 001 – R16 000	5 000	R1 600 – R1 999	400
R16 001 – R30 000	14 000	R2 000 – R2 499	500
Census 2001 & CS 2007 – Income	**Width**	R2 500 – R2 999	500
R1 – R400	400	R3 000 – R3 999	1 000
R401 – R800	400	R4 000 – R4 999	1 000
R801 – R1 600	800	R5 000 – R5 999	1 000
R1 601 – R3 200	1 600	R6 000 – R6 999	1 000
R3 201 – R6 400	3 200	R7 000 – R7 999	1 000
R6 401 – R12 800	6 400	R8 000 – R8 999	1 000
R12 801 – R25 600	12 800	R9 000 – R9 999	1 000
R25 601 – R51 200	25 600	R10 000 – R10 999	1 000
R51 201 – R102 400	51 200	R11 000 – R11 999	1 000
R102 401 – R204 800	102 400	R12 000 – R13 999	2 000
OHSs/LFSs/GHSs – Expenditure	**Width**	R14 000 – R15 999	2 000
R0 – R399	400	R16 000 – R19 999	4 000
R400 – R799	400	R20 000 – R24 999	5 000
R800 – R1 199	400	R25 000 – R29 999	5 000
R1 200 – R1 799	600	R30 000 – R39 999	10 000
R1 800 – R2 499	700	R40 000 – R49 999	10 000
R2 500 – R4 999	2 500		
R5 000 – R9 999	5 000		

For instance, if a household's exact monthly income and expenditure are both R8 200 in nominal terms, this household would fall in the 'R6 401 – R12 800' in CS 2007 (12 categories), 'R5 000 – R9 999' in GHS 2009 (10 categories) and 'R8 000 – R8 999' in AMPS 2009 (30 categories), and the derived income or expenditure amount (assuming the Pareto method is applied to the open interval and the midpoint method is applied to the other categories) would be estimated as R9 600, R7 500 and R8 500 respectively. In this case the AMPS amount (R8 500) is closest to the original amount (R8 200). Is the reliability of the derived amount being influenced by the number and width of bands in each survey? Would the poverty and inequality estimates be over-estimated or under-estimated as a result of these two factors?

There are no South African studies done to investigate the impact of the aforementioned issues on poverty and inequality estimates. Looking at international studies, Seiver (1979) found that income distribution results are influenced by the number and width of intervals chosen to span the range: fewer, wider brackets result in over-estimation of inequality measures. His study did not investigate the impact of the number and width of intervals on poverty.

Chapters 5 and 6 will return to this issue by investigating whether the extent of accuracy and reliability of income and expenditure information and the subsequent poverty and inequality estimates has any association with the number of intervals and width of the interval, when looking at the surveys that collect the above information in bands.

3.8 Households with zero or unspecified income

As discussion throughout Chapter 2, a serious problem of some surveys (especially the two censuses and CS 2007) is that a high proportion of people reported zero or unspecified personal income, which subsequently resulted in a large proportion of households with zero or unspecified household income (before hot deck imputation was applied by Stats SA in Census 2001). In addition, the proportion of households with unspecified household expenditure in the OHSs/LFSs and GHSs was also not low.

Regarding the people/households with missing personal/household income, Ardington et al. (2005) argue that if those with missing data fall excessively in the bottom of the income distribution, then poverty levels will be under-estimated if they are ignored. In contrast, if non-response is higher among the affluent, inequality measures are likely to be biased downwards[47]. Furthermore, with regard to the higher proportion of households with zero household income, even taking South Africa's high unemployment rates into consideration, it is highly unlikely that most of these zero-income households distinguished had no working-age members earning any income[48]. If these zero-income households are included for analyses, this could lead to an over-estimation of measured poverty and inequality.

With regard to missing data, there are three types of mechanisms, whereby (Lacerda et al., 2008: 6-9):

- Missing completely at random (MCAR): The distribution of missingness is independent of both the observed and missing data.
- Missing at random (MAR): The distribution of missingness is independent of missing data, but is dependent on some or all of the observed variables for each observational unit.
- Missing not at random (MNAR): The distribution of missingness is dependent on both the observed and missing data.

Hence, when analysing poverty and inequality, unless the data are MCAR, ignoring households with unspecified household income would lead to biased results. Including households that might incorrectly report zero income might lead to over-estimation of poverty and inequality levels. In general, the four main methods to deal with missing data are casewise deletion, available-case deletion, single imputation and multiple imputation. Each method is discussed in greater detail.

3.8.1 Casewise deletion

Casewise deletion, also commonly known as listwise deletion or complete-case analysis, is the simplest method to deal with missing data. It discards any observational unit with incomplete information (Lacerda et al. 2008: 11). Thus, in the case of household income data (or expenditure / consumption), those households that did not specify the household income amount or category (depending on how the question was asked) are immediately excluded from further analyses. However, as mentioned at the beginning of this section, if these

[47] Among households with unspecified household income, 35.2%, 29.4% and 47.6% contained at least one employed member, in Census 1996, Census 2001 and CS 2007 respectively. In addition, 27.9%, 22.2% and 37.1% of the heads of these households were employed at the time of each survey respectively. This implies that ignoring them would result in the over-estimation of poverty and narrowing of inequality.

[48] From Section 2.2.2.3, it was found that 13.0%, 21.0% and 8.2% of households in Census 1996, Census 2001 and CS 2007 had zero household income. When looking at these households in greater detail, 1.8%, 1.5% and 5.5% of them were headed by an employed member in each survey. 2.2%, 2.0% and 6.2% of these households had at least one employed member.

households are ignored, it would have a serious impact on the reliability of poverty and inequality estimates.

3.8.2 Available-case deletion

Available-case deletion is an extension of casewise deletion, but differs in that it only excludes those cases for which data are missing on the variables required to estimate the parameters of interest (Lacerda et al. 2008: 11). For example, if all households taking part in a survey reported dwelling type while 10% of households did not specify household income, but the latter variable is not used at all by a researcher in his/her analysis, then there is no need to worry about the missing income data, and all observations are kept in the dataset. However, if household income is an important variable for analysis (as in the case of this study), these 10% observations are immediately eliminated. However, excluding these households would have the same negative impact on poverty and inequality estimates as caused by casewise deletion. Thus, it seems the abovementioned two methods are not the best solution to deal with missing data for the purposes of this dissertation.

3.8.3 Single imputation

Imputation aims to provide reasonable estimates of the missing data, instead of simply ignoring observations with missing data. If it is applied to impute one value for each missing item of a variable, this is known as single imputation (Lacerda et al. 2008: 13). The commonly used single imputation methods are unconditional mean substitution, cell mean substitution, hot deck imputation, cold deck imputation and stochastic regression imputation.

<u>Unconditional mean substitution</u> means that the missing values are replaced by the mean of the observed values for that variable (Lacerda et al. 2008: 15). For example, assuming household income information from a survey was collected as exact amounts, 90% of households declared their household income and the mean household income for these households was R1 500. The household income of the 10% of households with unspecified income was assumed to be R1 500.

<u>Cell mean substitution</u> aims to divide respondents into cells on the basis of some known variables, and the average values within these cells are used for imputation (Malherbe 2007: 29 & Lacerda et al. 2008: 15). For example, the mean household income for a household headed by each race and gender could be derived. To apply this mean, a household headed by a black male has a mean household income of R1 600, then a household with exactly the same

race and gender characteristics but with unspecified household income is assumed to earn R1 600.

Hot deck imputation involves "substituting missing values with observed values drawn from similar responding units" (Lacerda et al. 2008: 16). For example, using the example above, households are divided into cells by race and gender of household head. After a random draw on a household headed by a white male, this household's income is R2 000. Then household A with unspecified household income but exactly the same race and gender characteristics has its household income imputed as R2 000. Similarly, after the second random draw on households from the same cell, a household with income level of R2 500 is chosen, and then household B with unspecified household income but the same race and gender characteristics has its household income imputed as R2 500. This process would carry on in each cell, until all missing household income data are imputed.

Cold deck imputation involves substituting missing values with a constant value from an external source (Lacerda at el. 2008: 16). For example, if a household headed by a black male taking part in IES 2000 did not answer the question "How much personal income tax did you pay the South African Revenue Service (SARS) in the last 12 months?", and from the National Treasury Budget Review 2000 document, it was found that, on average, a black male-headed household paid R1 500 personal income tax, then it would be assumed that the IES 2000 household as mentioned above spent R1 500 in the last 12 months to pay personal income tax to SARS.

Stochastic mean substitution is employed when "imputed values are randomly generated from a specified theoretical distribution with mean equivalent to the cell mean and variance equal to the cell variance" (Lacerda et al. 2008: 16). An extension to the above methods is known as stochastic regression imputation, in which "missing values are replaced by a value predicted by regression imputation plus a residual drawn to represent the uncertainty in the predicted value" (Lacerda et al. 2008: 17). For example, in the household income example above, in addition to race and gender of household head, other demographic characteristics such as the province of residence, age of household head, marital status of household head, as well as the number of children and elderly in the household should also be considered as explanatory variables to predict household income.

Finally, there are some less commonly used methods to deal with missing data. For example, the logical imputation method: A consistent value is estimated or deduced from other information relating to the individual or household, e.g., if two members from a household both declared they received old-age pension income in the last 12 months, but one of them stated he earned R1 500 from it while the other member did not specify his/her answer, then it is assumed that he/she also earned R1 500 from old-age pension during the same period. As another example, if both income and expenditure questions were asked in a household survey, but the respondent only declared the monthly household income as R10 000 but did not specify household expenditure, then one could impute the household expenditure as R10 000.

3.8.4 Multiple imputation

The multiple imputation method involves imputing several values for each missing item to allow for the inherent uncertainty in the imputation procedure. It consists of the following three steps (Lacerda et al. 2008: 17-18):

- m (which is greater than one – if m equals one, it stands for single imputation) plausible versions of the complete data are created by "imputing each missing value m times using m independent draws from an appropriate imputation model, conditional on the observed data" (Lacerda et al. 2008: 17);
- The m imputed datasets are then treated as if they are fully observed and analysed individually by standard complete-data methods;
- The results from the m analyses are combined in a single and proper manner so as to obtain overall estimates and standard errors that reflect both sample variation and uncertainty in association with the imputed values.

In this dissertation, values for the households with unspecified personal or household income are imputed using a particular multiple imputation technique developed by Raghunathan, Lepkowski, Van Howeyk and Solenberger (2001), which is applied when data are missing at random (MAR), namely sequential regression multiple imputation (SRMI). The SRMI method could be summarized as follows (Raghunathan et al. 2001: 86-87; Ardington et al. 2005: 8-11; Ardington, Lam, Leibbrandt and Welch 2006: 826-827; Lacerda et al. 2008; Vermaak 2008: 2-3):

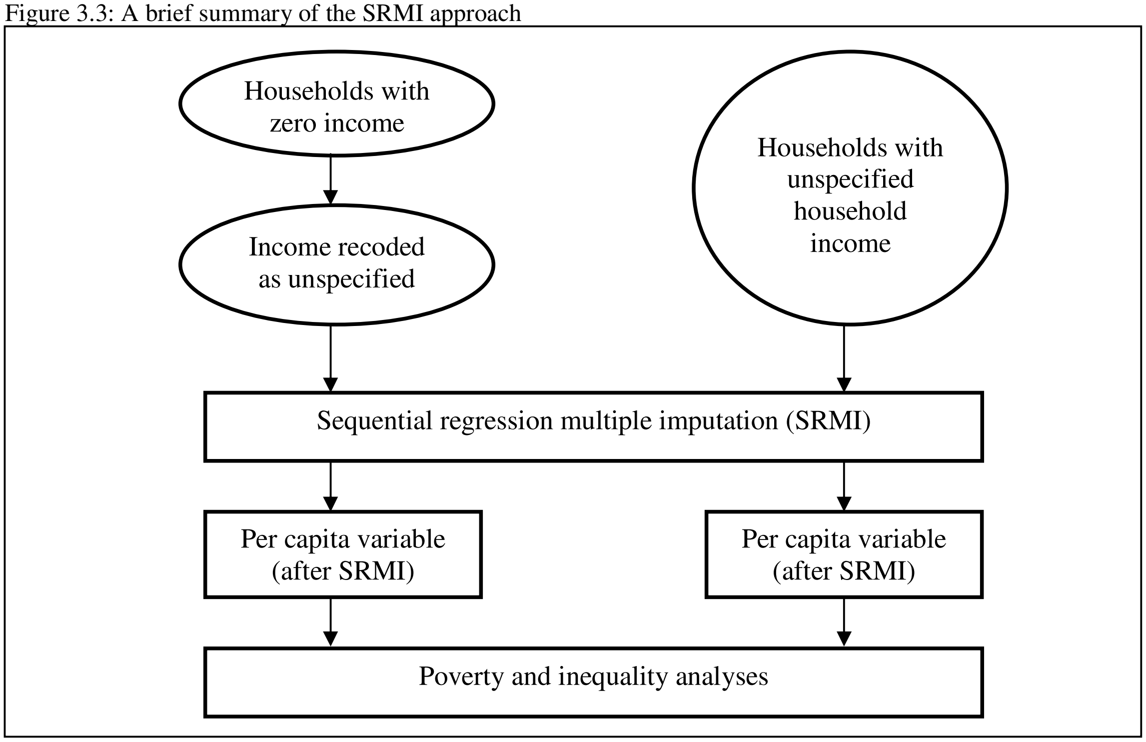
- The variables used in the imputation model are arranged from those with the least to those with the most missing values. The variables could be continuous (e.g., earnings amount), binary (e.g., gender), count (e.g., age), nominal categorical (e.g., province) or ordinal categorical (e.g., household income category).
- The matrix X represents all variables that are fully observed (i.e., there are no

unspecified responses), while $Y_1, Y_2, ..., Y_k$ stand for the ordered variables that contain missing values. The variables are arranged according to the extent of missing data they contain, i.e., Y_1 and Y_k have the least and most missing values respectively.

- All missing values are imputed as part of a process to estimate the joint conditional density of $Y_1, Y_2, ..., Y_k$ given X. In other words, $f(Y_1, Y_2, ..., Y_k | X, \beta_1, \beta_2, ..., \beta_k)$ $= f_1(Y_1 | X, \beta_1) f_2(Y_2 | X, Y_1, \beta_2) ... f_k(Y_k | X, Y_1, Y_2, ..., Y_k, \beta_k)$, where f_i and β_i stands for the conditional density functions and a vector of parameters in the conditional distribution respectively. In all cases, the β_i vectors are the estimated coefficients and estimates of the disturbance term.
- The first round of the imputation starts with Y_1 regressed on X in order to obtain an estimate of the β_1 vector. The missing values in Y_1 are then imputed by random draws from the predictive distribution. In other words, by first drawing a vector β_1^* from the posterior distribution of β_1 and then using β_1^* to generate a set of predicted values to replace the missing Y_1 values. A normal OLS regression model is used when Y_1 is a continuous variable. However, a Poisson model is used when Y_1 is a count variable, a logistic model is used when Y_1 is binary, a multinomial logistic model is used when Y_1 is a nominal categorical variable, and an ordered logistic model is used when Y_1 is an ordinal categorical variable.
- Since its missing values have now been imputed, Y_1 is appended to the set of predictor variables. Next, Y_2 is regressed on X and the newly derived Y_1 that includes the imputed values. The values are then imputed for Y_2. This imputation goes on until all Y variables have been imputed using non-missing variables (X) and all previously imputed variables of Y as covariates, before the first round is completed. At the end, the first complete set of data with no missing values is available.
- The imputation process is then repeated in the second round, updating the regression parameters β_i with parameters drawn from the now-complete distribution. That is, regress Y_1 on X and Y_2, Y_3, …, Y_k; regress Y_2 on X and Y_1, Y_3, …, Y_k; and so on. This cycle is repeated for a pre-specified number of rounds, or until the imputed values and parameters converge to a stable distribution.
- Assuming m stands for the number of imputations, m imputed complete datasets are produced at the end.

This SRMI approach is applied at both person and household levels to impute the household income or expenditure of missing data in this dissertation. Households with zero income or expenditure are recoded as missing, before SRMI is also applied on them (See Figure 3.3). For the remainder of the study, SRMI at person level and SRMI at household level will be referred to as SRMI1 and SRMI2 respectively. Chapter 5 will discuss the application of SRMI1 and SRMI2 on the survey data in greater detail, and Chapter 6 will examine how the poverty and inequality estimates would change before and after the application of SRMI to deal with households with zero or unspecified income or expenditure.

Figure 3.3: A brief summary of the SRMI approach



3.9 External validation to improve the reliability of survey data

Survey data should be validated against various external sources in order to determine the reliability of it. These sources are discussed in this section.

3.9.1 Validation against national accounts

As discussed throughout Chapter 3, surveys are more likely to under-estimate income / expenditure / consumption than to over-estimate it, due to reasons like fatigue, loss of interest, lack of motivation, illiteracy, recall bias, telescoping, and the tendency to declare zero or unspecified income, even if the households contain members who are employed or have income support from non-labour sources. Hence, one view is that the distributional estimates of the survey data should be adjusted rightwards to be consistent with the national accounts series for aggregate household income / consumption (Van der Berg, Burger, Burger, Burger, Louw and Yu 2005 & 2009). That is, household survey means are replaced by national accounts means, but the distribution of the household survey is retained.

However, adjusting survey means in line with national accounts mean implies the following must be true (Deaton 2001: 135): (1) the national accounts estimates are correct; (2) survey estimates of the mean are incorrect; (3) in spite of (2), the income / consumption levels of each household in the survey are correct up to a multiplicative factor. Proponents of the adjustment procedure generally believe that national accounts data are, in general, superior to survey data, and argue that not adjusting the survey means is more likely to introduce a larger error into the trends than adjusting the means[49]. However, if the sources of data disagree and there is no reason to favour one over the other, a more modest version of adjustment is suggested, that is, the survey data are scaled up by some weighted average of the national accounts mean and the survey mean, at least after correcting for conceptual differences and coverage (Deaton 2001: 136). The possible problems of national accounts data as well as reasons why adjusting the survey means might even create more negative effects on the reliability of poverty and inequality estimates are the focus of this section.

First, it is argued by some (Ravallion 2000; Deaton 2001: 133-134; Karshenas 2003: 694; Ravallion 2003: 646) that the national accounts estimates of consumption might not be the ideal variable to be treated as the gold standard to which the survey estimates should correspond. While the consumption measure in household survey is derived from self-reported expenditures (e.g., cash and from own stock) by the households in the interviews, households are treated as residual claimants in the national accounts, as aggregate consumption is simply the residual obtained by subtracting other measured forms of domestic absorption from aggregate output. Hence, the errors and omissions in the estimation of the other components of the gross domestic product (GDP) all impinge on aggregate consumption.

The second problem with the national account estimates of consumption is that they implicitly include spending by unincorporated businesses and non-profit organizations, for example, religious groups, trade unions, clubs, and political parties. However, these estimates are not

[49] An example is the rapid decline of income and expenditure between IES 1995 and IES 2000. The magnitude of the measured decline is even greater than the fall in output during the Great Depression. Hence, the poverty rate would show a rapid decline between the two surveys, and such decrease would be smaller had the distribution of the 2000 data been adjusted in line with national accounts mean.

captured in surveys, as the aforementioned institutions are not households and hence did not take part in the surveys. Hence, the growth measured in the national accounts consumption might not really show up in progresses in the living standards of the poor, and if the survey income / consumption distribution is adjusted (rightwards) in line with the national accounts consumption mean, this would result in an under-estimation of poverty (Ravallion 2000; Deaton 2001: 133-134; Karshenas 2003: 694; Ravallion 2003: 646-647).

Thirdly, Ravallion (2000 & 2003: 646-647) and Deaton (2001: 133-134 & 2005: 10) argue that rich households are missed more than the poor by surveys (i.e., unit non-response takes place), as the well-off households are more likely to refuse to participate in the survey, or it is relatively more difficult to penetrate the gated communities (e.g., getting past the guard dogs) in which many rich people live. Hence, such households could be replaced by the more compliant but perhaps less well-off ones. Furthermore, even if the rich households take part in the survey, the included rich people are more likely to understate their income / consumption more than the included poor do, and this implies that inequality is under-estimated.

If the survey mean is simply replaced by the national accounts mean, it assumes that the survey under-estimates income / consumption by a constant proportion across all levels. Thus, if this were untrue, after the adjustment, the income / consumption of the poor households could be seriously over-estimated, and poverty would in turn be under-estimated. As an example, the bottom 20% and top 20% of the population under-stated their expenditures by 25% and 50% respectively, while the average household under-stated its expenditure by 35% (when comparing with national accounts mean), if there is a uniform rightward adjustment of the survey mean in line with the national accounts mean by 35%, this clearly results in the over-estimation of expenditure of the poor households, and a subsequent under-estimation of poverty. This implies that the simple adjustment of the survey distribution upwards in line with the national accounts mean might not help improving the survey poverty and inequality estimates, if the unreliable survey distribution is the root of the problem but is not addressed.

It might also be argued that surveys have missed the poor rural households (as it is expensive or dangerous to visit these places) as well as the very poor without fixed abode (i.e., homeless), and as a result of failing to include these poor households to take part in the survey, the survey income / consumption estimates would be biased upwards. Hence, once again, the main problem has to do with the incorrect distribution of survey data as a result of failing to capture these poor households as part of the sample, and simply adjusting the survey mean in line with national accounts by assuming the extent of adjustment is uniform across the whole population might not improve the reliability of poverty and inequality estimates, but rather complicate matters.

Based on the above arguments, different kinds of households have different likelihoods of being included in household surveys. As a result, survey results need to be weighted correctly to give an accurate representation of the population as a whole, with the calculation of suitable weights depending on the availability of accurate, up-to-date information about the population (Deaton 2001: 133-134). This implies that the replacement of survey means by national accounts means does not improve the poverty and inequality estimates at all, and might even worsen them, if the issues relating to the survey weights are not sorted out right at the beginning.

Other problems affecting the comparability between national accounts and household survey estimates are related to the capture of informal economic activities and certain income items. First, Deaton (2005) and Ravallion (2003: 646-647) argue that the value of informal activities is notoriously difficult to measure in the national accounts. Hence, as an economy grows and its structures change, many production activities shift from the informal sector to the formal sector. Consequently, economic activity is increasingly accurately captured in the national accounts data. This implies that the level of national accounts income is understated but growth is overstated as the economy develops and grows. This could partly explain the diverging gap between national accounts and household survey estimates of income in countries like India[50] (Deaton and Kozel 2005). Secondly, in the national accounts income and private consumption estimates, items like imputed rent and in-kind income are taken account of, but they might not be recorded in household surveys, and this eventually results in the differences between the two series[51].

Chapter 5 will revert back to this issue by comparing the survey income and expenditure with the national accounts income of the same year to determine if the surveys poorly captured the information. Moreover, Chapter 6 will examine whether the poverty levels and trends would differ, had the survey income and expenditure distributions been adjusted in line with the

[50] Whether this diverging gap also took place in the South African survey data is investigated in Chapter 5.

[51] As discussed in Chapter 2, IES 2005/2006 and NIDS are two surveys containing questions that clearly asked the respondents to declare imputed rent and in-kind income, and these items were taken into consideration when household income and consumption were derived. This is not the case in other surveys under study, as respondents were simply asked to declare income or expenditure from all sources, but some respondents might not be aware that imputed rent and in-kind income are income or expenditure items. The impact of the inclusion of the imputed rent on poverty and inequality estimates is studied in Chapter 6.

national accounts income mean.

3.9.2 Validation against other external sources

In addition to the national accounts, the survey data could also be validated against other external sources. Some of the commonly chosen external sources are discussed here. The focus is on the validation of IES data against these sources. First, the survey data on social grants income could be compared with the social grants expenditure by the National Treasury. For example, Table 3.3 below shows that, in general, the old-age or war pension income as captured by IES 2000 and IES 2005/2006 was higher than the amounts as reported by the National Treasury (see the percentages like 132.1%, 126.2%, 136.5% and 126.3% in the table), while disability grant income was under-captured the IESs by between 20% and 30%.

Table 3.3: Social grants income of IES 2000 and 2005/2006 compared with social grants expenditure of National Treasury (Rand million, nominal terms)

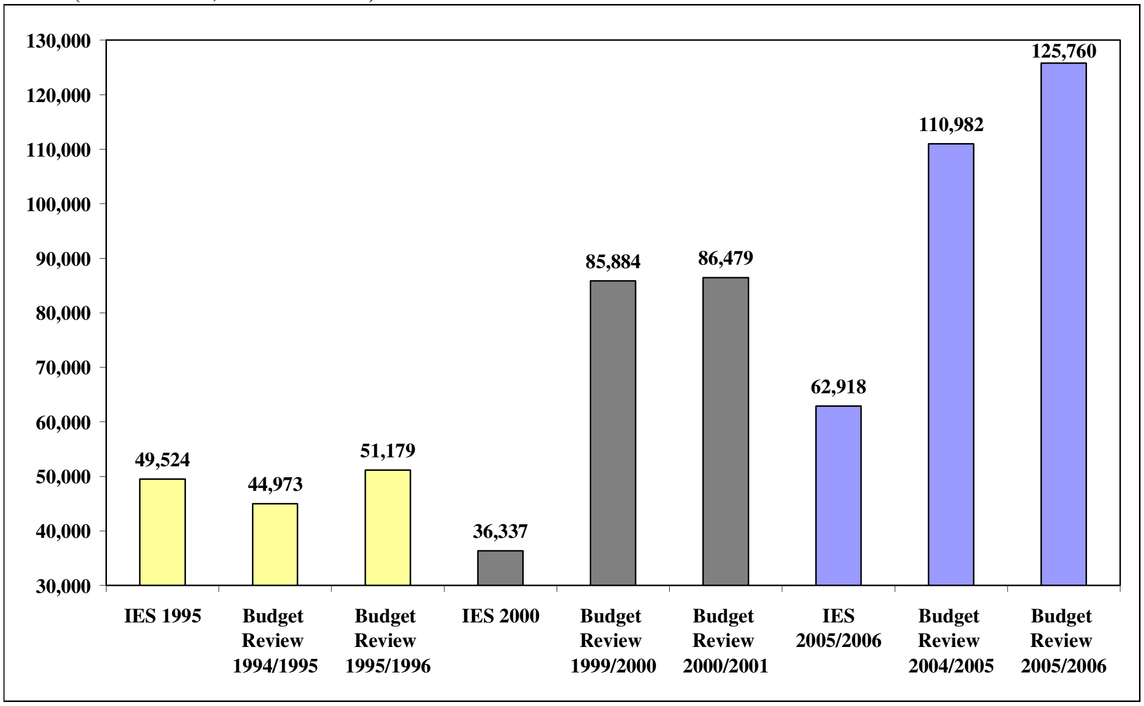
	Old-age/War pension	**Disability grant**	**Child/Family/Other grants**
[A]: IES 2000	15 402	3 058	1 533
[B]: Treasury - 1999/2000	11 660	3 823	944
[C]: Treasury - 2000/2001	12 208	4 066	1 770
[A] / [B]	132.1%	80.0%	162.4%
[A] / [C]	126.2%	75.2%	86.6%
[D]: IES 2005/2006	25 301	10 375	19 981
[E]: Treasury - 2004/2005	18 540	12 570	13 774
[F]: Treasury - 2005/2006	20 025	14 438	17 465
[D] / [E]	136.5%	82.5%	145.1%
[D] / [F]	126.3%	71.9%	114.4%

Data sources: Own calculations using IES data and National Treasury Budget Review (various issues).

Secondly, net personal income tax expenditure data of the survey could be compared with the net personal income tax revenue received by SARS. From Figure 3.4, it can be seen that IES 1995 did an outstanding job of capturing this tax expenditure accurately. However, the income tax expenditure captured in IES 2000 is only equivalent to slightly above 40% of the income tax revenue of SARS in both the 1999/2000 and 2000/2001 budget. This under-estimation of tax expenditure in IES 2000 could be associated with the very low total income captured in the survey compared with the national accounts total income in the same year[52]. The under-capture of income tax expenditure also took place in IES 2005/2006, despite the extent of it being less serious (about 57% of the income tax revenue of SARS as reported in the 2004/2005 and 2005/2006 budget was captured).

[52] This will be discussed in greater detail in Chapter 5.

Figure 3.4: Net personal income tax expenditure of IESs compared with net personal income tax revenue of SARS (Rand million, nominal terms)

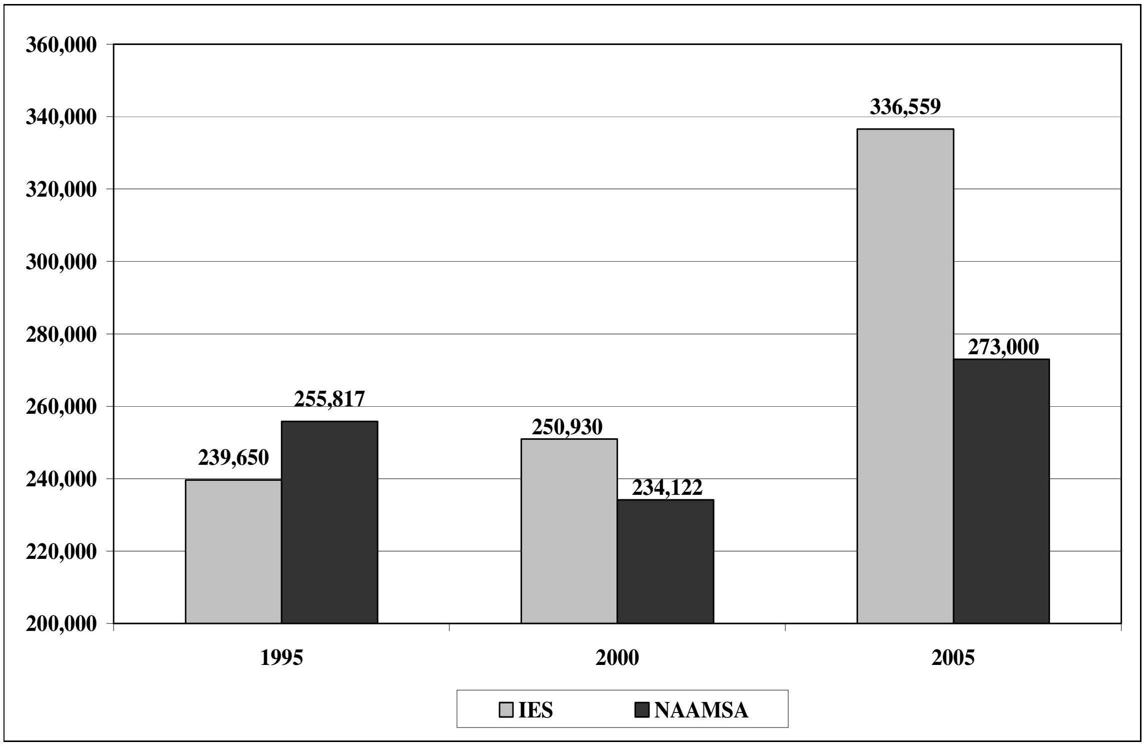


Data sources: Own calculations using IES data and National Treasury Budget Review (various issues).

In the three IESs, household heads were asked to declare expenditure on new and used vehicles. Thus, the statistics on the number of new cars sold from the National Association of Automobile Manufacturers of South Africa (NAAMSA) could be compared with the number of households with non-zero expenditure on new and used vehicles in the IESs. A drawback of the latter data is that it is impossible to know the number of new vehicles purchased in each household, and hence, the IES and NAAMSA data could only be compared based on the assumption that each household reporting non-zero new vehicle spending in the IESs only purchased one new vehicle.

The results from Figure 3.5 show that both IES 2000 and IES 2005/2006 over-estimated the number of new motor vehicle purchases when compared with NAAMSA data, but the extent of over-estimation was greater in the latter survey.

Figure 3.5: Number of households with non-zero expenditure on new vehicle purchase in IESs compared with number of new vehicles sold from NAAMSA data



Data sources: Own calculations using IES and NAAMSA data.

Finally, the survey data on petrol expenditure could be compared with the estimated petrol cost released by the South African Petroleum Industry Association (SAPIA). For instance, Table 3.4 compares the estimated total cost of petrol as reported by SAPIA and the total petrol expenditure from the IESs, and the results show that petrol expenditure in IES 2000 and 2005/2006, as for the personal income tax expenditure, was seriously under-estimated, when compared with external sources. The under-estimation was more seriously in IES 2005/2006. These results are quite similar to what happened to what happened in the net personal income tax expenditure data in the IESs (i.e., under-estimation took place in both cases).

Table 3.4: Petrol expenditure in the IESs compared with estimated petrol cost from SAPIA

IES	[A]: IES petrol expenditure (Rand million)	[B]: SAPIA (million litre)	[C]: SAPIA: Fuel price per litre (97, Coast)	[D] = [B] × [C] Estimated total cost (Rand million)	[A] / [D]
1995	R7 277	10 020	0.5708	R5 720	127%
2000	R12 852	10 556	1.9511	R20 593	63%
2005/2006	R23 533	11 158	4.9527	R55 263	43%

Data sources: Own calculations using IES and SAPIA data.

Note: The IES 1995 data are compared with the aggregate of SAPIA's 1994Q4, 1995Q1, 1995Q2 and 1995Q3 data, the IES 2000 data are compared with the aggregate of SAPIA's 1999Q4, 2000Q1, 2000Q2 and 2000Q3 data, and the IES 2005/2006 data are compared with the sum of SAPIA's 2005Q4, 2006Q1, 2006Q2 and 2006Q3 data.

3.10 Post-stratification weighting

As discussed in Chapter 2, with the exception of Census 1996 and Census 2001, the remaining data used for the labour market, poverty and inequality analyses in this study is survey data, as only a sample of people from the population took part in the survey. Design weights are created to make the sample represent the population. Different households have different inclusion probabilities as a result of both designed and unplanned factors. Hence, some households are over-represented relative to the others, and vice versa. In order for the sample estimates to accurately reflect the population, there is a need to weight each household according to its true inclusion probability.

In addition, due to the presence of non-coverage and unit non-response, post-stratification adjustment to the design weights is necessary by benchmarking the survey data to external aggregate population data so as to impose consistency between survey results and those from external sources. In the Stats SA survey data under study (IESs, OHSs/LFSs/QLFSs and GHSs), the person weights were post-stratified to the external population totals, i.e., the mid-year population estimates at the time of the survey derived by using the Census 1991, 1996 and 2001 information, with the pre- and post-census year's population information being calculated using exponential interpolation and extrapolation.

Nonetheless, some concerns were raised regarding the reliability of the post-stratification design weights (Branson 2009):

- The auxiliary data (i.e., the mid-year population estimates) used as a benchmark in the post-stratification adjustment could be unreliable, inconsistent over time and of poor quality, thereby resulting in temporal inconsistencies even at the aggregate level. Branson (2009: 14) argues that this is likely the case in the population data derived by the Census, as the data are outdated to be used to project population estimates over a long period. Hence, the increased precision of the post-stratification weights could be offset by the potential bias introduced by using the questionable auxiliary data;
- Since the survey data are cross sectional, the purpose of the post-stratification adjustment is to produce the best estimates of the population, given the information available at the time of the survey. However, temporal consistency is not considered. This creates problems when the data are used for time-series analyses;
- As the post-stratification adjustment of the Stats SA data was conducted at the person level (i.e., the person weight), this could result in inconsistency between person-level

and household-level data, and the resultant analyses done at person and household levels would not necessarily agree.

Table 3.5 presents the racial, gender, provincial and urban shares of the population in the surveys under study. It can be seen that these shares are quite stable and comparable across the surveys, with a few exceptions. First, the white share has always been about two percentage points higher (at the expense of the black share) in AMPSs, when compared with the OHSs, LFSs and GHSs of the same years. In addition, the male share in AMPSs is also about two percentage points higher. Furthermore, with regard to the proportion of the population residing in urban areas at the time of the survey, PSLSD and OHS 1994 are the only surveys with this share below 50% (44.51% and 48.95% respectively), while this share is much higher in NIDS (59.7%) when compared with the urban shares of other surveys. These findings could be attributed to different weighting techniques applied in each survey.

Hence, the entropy post-stratification approach is adopted to re-weight the person weights of all the data under study, with the person weights being adjusted to conform to the race, gender and age distribution of the population estimates as calculated by the Actuarial Society of South Africa 2003 (ASSA 2003) model. Branson (2009: 17) argues that the population data derived from the ASSA model is more time consistent.

The ASSA 2003 model aims to project the South African mid-year population from 1985, on the basis of various demographic, epidemiological and behavioural assumptions. The model could also be used to project trends in fertility and mortality as well as HIV/AIDS prevalence rate. There were two ASSA 2003 models at the time of this study: the full model projects the population of the four race groups by gender and age category (18 categories in total: 0-4 years, 5-9 years, and so forth, with the last category being “85 years or above) as well as the provincial population, while the lite model does not divide the population by race.

The entropy approach could be explained as follows: let x be a random variable with possible outcomes $x_k, k = 1,2,...,K$ and probabilities, $p = (p_1, p_2,..., p_k)'$, then the entropy measure is: $H(p) = -\sum_k p_k \ln p_k$, where $0 \cdot \ln(0)$ is defined to be 0. H(p) = 0 presents the degenerate solution, one possible outcome with certainty. H(p) reaches a maximum when the probability distribution is uniform. This is referred to as the maximum entropy (ME) approach.

Table 3.5: Racial, gender, provincial and urban shares of total population, selected surveys

	Total population	**Racial share**				**Male share**	**Provincial share**									**Urban share**
		Black	**Coloured**	**Indian**	**White**		**WC**	**EC**	**NC**	**FS**	**KZN**	**NW**	**GAU**	**MPU**	**LIM**	
Census 1996	37,341,483	76.5%	9.1%	2.8%	11.7%	48.01%	10.02%	14.88%	2.06%	6.53%	20.67%	8.31%	17.94%	7.08%	12.51%	53.50%
Census 2001	41,747,214	79.3%	8.9%	2.6%	9.2%	47.91%	9.93%	14.55%	1.83%	6.21%	20.91%	8.19%	19.73%	6.89%	11.76%	56.44%
CS 2007	47,374,601	79.3%	8.5%	2.5%	9.7%	48.28%	10.60%	13.90%	1.84%	5.70%	20.42%	7.57%	21.62%	7.21%	11.15%	N/A[#]
IES 1995	41,530,563	76.3%	8.6%	2.5%	12.6%	48.02%	9.77%	15.55%	2.07%	6.48%	20.75%	8.25%	18.07%	6.89%	12.17%	50.31%
IES 2000	43,285,432	78.5%	8.8%	2.6%	10.1%	49.19%	9.42%	15.50%	1.95%	6.48%	20.74%	8.25%	18.48%	6.94%	12.24%	53.33%
IES 2005	47,380,368	79.5%	8.8%	2.5%	9.2%	47.89%	10.01%	14.46%	2.40%	6.24%	20.95%	6.96%	20.22%	7.41%	11.34%	58.98%
OHS 1994	40,251,142	75.9%	8.6%	2.6%	12.9%	49.05%	9.12%	15.92%	1.83%	6.40%	21.28%	8.00%	17.18%	7.19%	13.08%	48.95%
OHS 1995	39,659,831	77.1%	9.1%	2.6%	11.2%	48.02%	9.77%	15.55%	2.07%	6.48%	20.75%	8.25%	18.07%	6.89%	12.17%	50.31%
OHS 1996	40,582,538	77.3%	9.1%	2.5%	11.1%	48.06%	9.75%	15.54%	2.07%	6.49%	20.74%	8.27%	18.10%	6.90%	12.15%	53.66%
OHS 1997	41,443,101	77.4%	9.1%	2.6%	10.9%	48.19%	9.72%	15.53%	2.07%	6.49%	20.75%	8.28%	18.12%	6.91%	12.13%	54.18%
OHS 1998	42,212,335	77.8%	9.0%	2.5%	10.7%	48.27%	9.67%	15.60%	2.05%	6.48%	20.69%	8.30%	18.00%	6.92%	12.29%	54.08%
OHS 1999	43,271,686	78.0%	8.9%	2.6%	10.5%	48.38%	9.63%	15.63%	2.05%	6.49%	20.78%	8.29%	17.96%	6.92%	12.26%	53.91%
LFS 2000b	44,821,345	78.6%	8.8%	2.5%	10.1%	49.19%	9.42%	15.50%	1.95%	6.48%	20.74%	8.25%	18.48%	6.94%	12.24%	53.33%
LFS 2001b	45,081,045	78.7%	8.8%	2.5%	10.0%	49.19%	9.44%	15.48%	1.94%	6.46%	20.73%	8.24%	18.52%	6.94%	12.24%	53.15%
LFS 2002b	45,560,990	78.9%	8.8%	2.5%	9.8%	49.18%	9.57%	15.36%	1.94%	6.42%	20.68%	8.22%	18.70%	6.92%	12.18%	53.58%
LFS 2003b	46,046,026	79.1%	8.8%	2.5%	9.6%	49.19%	9.69%	15.23%	1.93%	6.38%	20.66%	8.20%	18.90%	6.90%	12.12%	53.29%
LFS 2004b	46,490,122	79.2%	8.8%	2.5%	9.5%	49.20%	9.84%	15.08%	1.93%	6.33%	20.60%	8.12%	19.17%	6.88%	12.05%	N/A[#]
LFS 2005b	46,917,195	79.4%	8.8%	2.5%	9.3%	49.21%	9.93%	14.99%	1.92%	6.29%	20.57%	8.15%	19.29%	6.86%	12.01%	N/A[#]
LFS 2006b	47,429,106	79.5%	8.8%	2.4%	9.2%	49.23%	10.03%	14.53%	2.30%	6.23%	20.93%	7.11%	20.14%	7.40%	11.32%	N/A[#]
LFS 2007b	47,882,965	79.6%	8.9%	2.4%	9.1%	49.25%	10.14%	14.41%	2.30%	6.19%	20.91%	7.09%	20.29%	7.39%	11.29%	N/A[#]
QLFS 2008Q4	48,780,039	79.2%	9.0%	2.6%	9.2%	48.17%	10.83%	13.48%	2.30%	5.90%	20.74%	7.03%	21.52%	7.37%	10.83%	N/A[#]
QLFS 2009Q4	49,148,442	79.3%	9.0%	2.6%	9.1%	48.21%	10.89%	13.39%	2.29%	5.87%	20.69%	7.01%	21.66%	7.36%	10.83%	N/A[#]
GHS 2002	45,402,018	79.3%	8.8%	2.5%	9.4%	47.79%	10.15%	14.26%	1.80%	5.98%	20.97%	8.19%	19.98%	6.99%	11.69%	55.25%
GHS 2003	46,445,749	79.6%	8.9%	2.4%	9.1%	47.76%	10.23%	13.99%	1.76%	5.89%	21.01%	8.17%	20.32%	6.99%	11.64%	54.89%
GHS 2004	46,410,515	79.2%	8.8%	2.5%	9.5%	49.19%	9.79%	15.13%	1.93%	6.34%	20.63%	8.18%	19.05%	6.89%	12.07%	53.60%
GHS 2005	46,858,740	79.4%	8.8%	2.5%	9.3%	49.21%	9.91%	15.01%	1.92%	6.29%	20.57%	8.15%	19.26%	6.86%	12.02%	N/A[#]
GHS 2006	47,337,997	79.5%	8.8%	2.5%	9.2%	49.23%	10.01%	14.88%	1.92%	6.24%	20.53%	8.14%	19.45%	6.86%	11.97%	N/A[#]
GHS 2007	47,796,008	79.6%	8.9%	2.4%	9.1%	49.25%	10.12%	14.44%	2.30%	6.20%	20.92%	7.09%	20.26%	7.39%	11.29%	N/A[#]
GHS 2008	48,640,831	79.2%	9.0%	2.6%	9.2%	48.16%	10.81%	13.50%	2.31%	5.90%	20.72%	7.04%	21.53%	7.36%	10.83%	N/A[#]
GHS 2009	49,334,589	79.3%	9.0%	2.6%	9.1%	48.40%	10.87%	13.47%	2.32%	5.88%	21.19%	6.99%	21.38%	7.31%	10.59%	N/A[#]
PSLSD 1993	41,476,952	78.5%	7.8%	2.5%	11.2%	48.46%	10.96%	13.35%	2.31%	5.85%	21.16%	6.98%	21.57%	7.29%	10.53%	44.51%
NIDS 2008	48,644,588	79.3%	8.8%	2.5%	9.3%	48.23%	10.80%	13.51%	2.32%	5.91%	20.77%	7.04%	21.46%	7.37%	10.80%	59.70%
AMPS 1997	39,824,647	76.7%	8.2%	2.3%	12.8%	46.47%	9.63%	14.50%	1.82%	6.02%	21.37%	7.79%	19.11%	7.64%	12.13%	54.54%
AMPS 2000	44,052,224	77.6%	8.2%	2.2%	12.0%	47.10%	9.08%	14.85%	1.92%	6.09%	23.24%	8.14%	18.50%	7.05%	11.14%	56.22%
AMPS 2004	45,112,432	78.0%	8.4%	2.3%	11.3%	47.48%	9.47%	15.36%	1.86%	5.98%	22.09%	8.41%	17.52%	7.05%	12.26%	55.28%

The area type variable is not available in CS 2007, LFS 2004b-LFS 2007b, all QLFSs and GHS 2005-2009.

The maximum entropy approach can be generalized to include prior information about the probability distribution with the aim to improve the accuracy of the estimates. This is known as the cross entropy (CE) approach and could be explained as follows: consider a survey sample of K individuals prior to adjustment probabilities q_k, i.e., the initial Stats SA person weights converted into proportions to the sum of one. Each individual has a vector of x_k characteristics (e.g., race, gender, age group). The CE estimate of p is the estimate which minimizes the difference from q, given the constraints to the problem (to be explained below). Alternatively, this implies the person weights are adjusted to meet aggregate trends (as derived by the ASSA model) which appear realistic over time, while simultaneously diverging as little as possible from the original Stats SA person weights.

In equation terms, the CE approach could be explained as follows (Golan, Judge and Miller 1996; Branson 2009: 34-36):

$\underset{p_k}{Min}\, I(p,q) = \underset{p_k}{Min}\left(\sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{q_k}\right)\right) = \underset{p_k}{Min}\left(\sum_{k=1}^{K} p_k \ln p_k - \sum_{k=1}^{K} p_k \ln q_k\right)$, subject to the moment consistency constraints $\sum_{k=1}^{K} p_k x_t = y_t \quad t \in [1,\ldots,T]$ and adding-up normalization constraint $\sum_{k=1}^{K} p_k = 1$.

Each x_t stands for a person-level indicator, indicating which demographic group the individual is in (e.g., the individual's gender, age category and race). T represents the number of restrictions. For example, if race (4 categories), gender (2 categories) and age groups (18 categories) are used, altogether there are 144 race-gender-age constraints (4 × 2 × 18), nine provincial constraints, plus the category "missing" (i.e., those with unspecified race, gender or age), i.e., 154 (144 + 9 + 1) constraints in total.

The new probability person weights are estimated as follows:

$$\underset{p_k}{Min}\, L = \underset{p_k}{Min}\left(\sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{q_k}\right) + \sum_{t=1}^{T} \lambda_t\left(y_t - \sum_{k=1}^{K} p_k x_k\right) + \mu\left(1 - \sum_{k=1}^{K} p_k\right)\right)$$

The first-order conditions are:

$$\frac{\partial L}{\partial p_k} = \ln p_k - \ln q_k + 1 - \sum_{t=1}^{T} \lambda_t x_t - \mu = 0 \qquad k \in [1,...,K]$$

$$\frac{\partial L}{\partial \lambda_t} = y_t - \sum_{k=1}^{K} p_k x_k = 0 \qquad t \in [1,...,T]$$

$$\frac{\partial L}{\partial \mu} = 1 - \sum_{k=1}^{K} p_k = 0$$

The solution to this can be written as:

$$p_k = \frac{q_k}{\Omega\left(\tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_T\right)} \exp\left[\sum_{t=1}^{T} \tilde{\lambda}_t x_k\right] \qquad k \in [1,...,K], \text{ where } \Omega\left(\tilde{\lambda}\right) = \sum_{k=1}^{K} q_k \exp\left[\sum_{t=1}^{T} \tilde{\lambda}_t x_k\right]$$

Once the entropy person weights are derived, the household entropy weight variable is created and is equal to the mean entropy person weight within the household. The CE weights will be later used to investigate the labour market, poverty and inequality trends, with their results compared to those obtained by using the original person and household weights.

The most efficient way to adjust the person weights would be to use the original design person weights (i.e., before the post-stratification adjustment against the Census mid-year population estimates). However, these weights are not publicly available and hence the adjusted design person weights (i.e., after the adjustment against the Census estimates) are used. These weights will be re-weighted against the ASSA 2003 model's population estimates in the CE approach before the labour market trends as well as the poverty and inequality trends are re-examined in Chapters 4 and 6 respectively.

The approach discussed above was adopted by Branson (2009), the only South African study that investigated the labour market trends using the entropy approach (her study did not analyse the poverty and inequality trends). The person weights of OHS 1995-1999 and the March LFSs in 2000-2004[53] were re-weighted. After that, Branson looked at the trends in the share of single-person households, population shares by gender and area type of residence respectively, economically population and the number of employed, by using the Stats SA person weights as they were, the adjusted person weights after ME approach and the adjusted

[53] When imposing the ASSA 2003 model's population estimates constraints on the entropy model, Branson (2009) combined the "80-84 years" and "85 years or above" categories together as "80 years or above". In other words, there were 17 age categories in total. Altogether there are 136 race-gender-age constraints (4 × 2 × 17), 9 provincial constraints, plus the category "missing" (i.e., those with unspecified race, gender or age), i.e., 146 (136 + 9 + 1) constraints in total.

person weights after the CE approach.

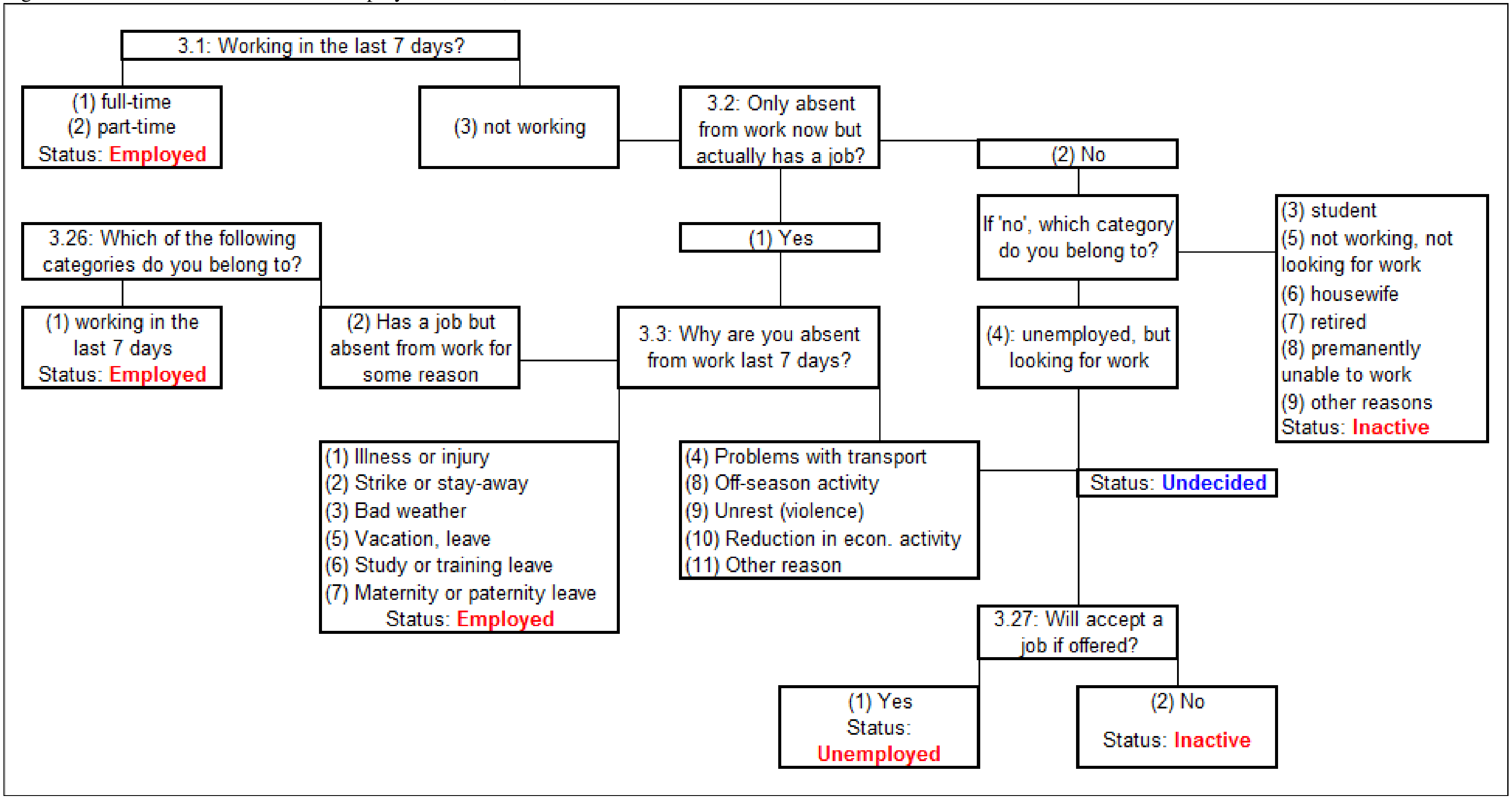
In particular, she investigated whether the abrupt changes during certain years (especially in the OHSs and the changeover from OHS to LFS) were attributed to the inappropriate post-stratification technique by Stats SA or rather due to other reasons like changes in the questionnaire design, etc. After the entropy approach was adopted, it was found that "although there are small changes, the entropy weights have no significant effect in creating a more consistent trend in the labour market variables between 1995 and 2004. In other words, the large inconsistencies in the labour market variables are not a result of shifts in the weights" (Branson 2009: 53). The same findings were observed regardless of whether the ME or CE approach was conducted. Furthermore, Branson (2009: 53) found that the relatively higher employment levels in OHS 1995 (compared with OHS 1996-1997) and LFS 2000a (a rapid 1.5 million increase from the OHS 1999 employment level) were "unlikely to be a function of incorrect weights caused by post-stratification errors", but these abrupt changes were rather "either real or the result of measurement error".

Chapter 4 will extend Branson's analyses by re-weighting all OHSs, LFSs and QLFSs before re-examining the labour market estimates and trends to determine whether the abrupt changes in the labour market aggregates are real, due to the different sampling techniques adopted by Stats SA, or other factors (e.g., the changes in the methodology to derive labour market status of the respondents, to be discussed in Section 3.11). Furthermore, in Chapter 6, the re-weighting approach would be conducted in all surveys under study, if possible, before investigating if there are any significant differences in the poverty and inequality estimates and trends after the datasets are re-weighted.

3.11 Labour market status derivation methodology

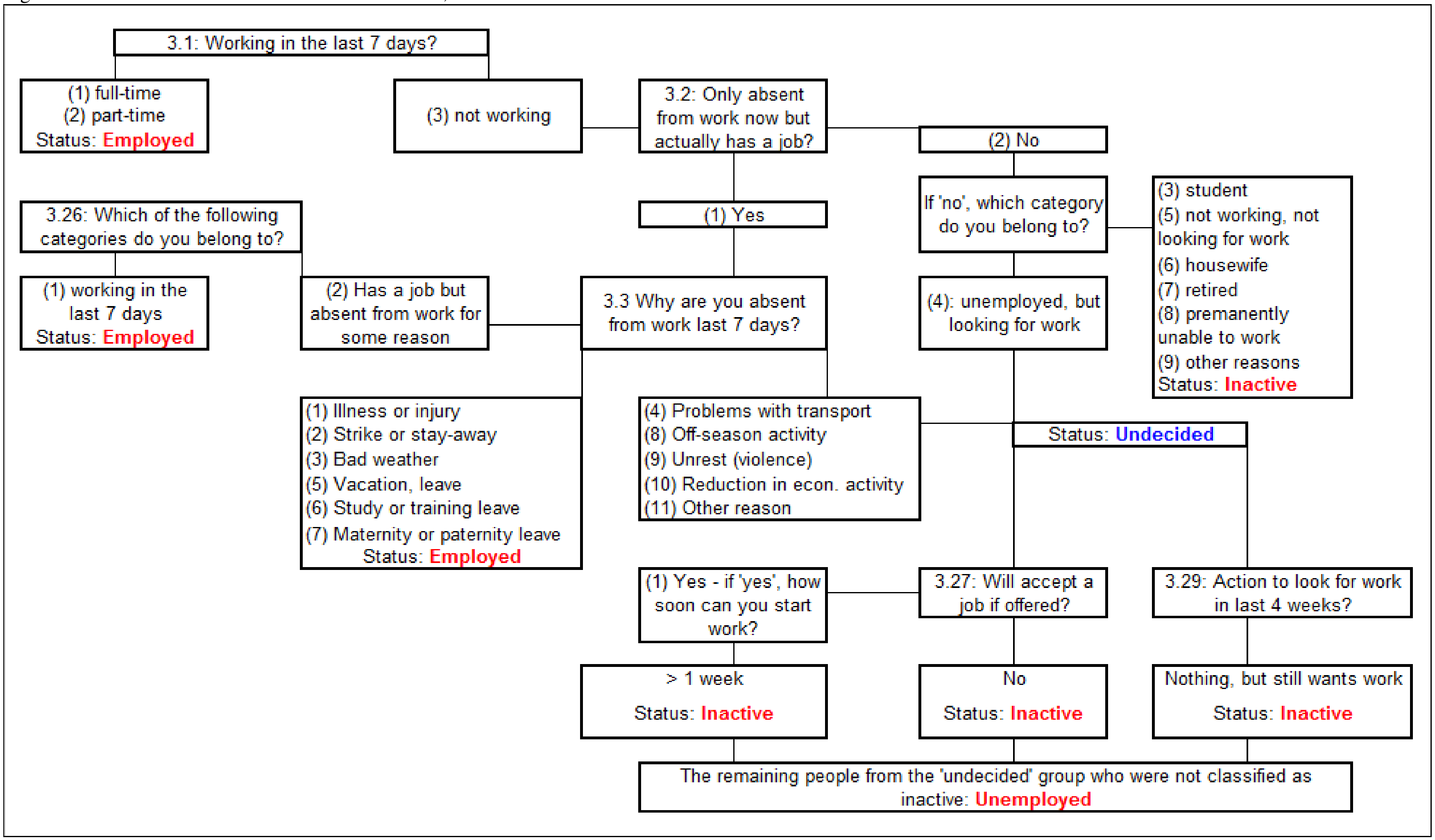
As mentioned in Section 2.4, OHSs, LFSs and QLFSs are the major sources of labour market information, and it was only briefly discussed that the narrow and broad labour market status of the respondents could be derived. This section will discuss the changes in the methodology to derive labour market status of respondents in the OHSs, LFSs and QLFSs in greater detail. Figures 3.6 to 3.15 present the detailed algorithms on how the narrow and broad labour market status of the participants was derived in each survey. As mentioned before, the OHS 1994-1995 methdogologies are not provided by Stats SA.

Figure 3.6: Derivation of broad labour employment status, OHS 1996



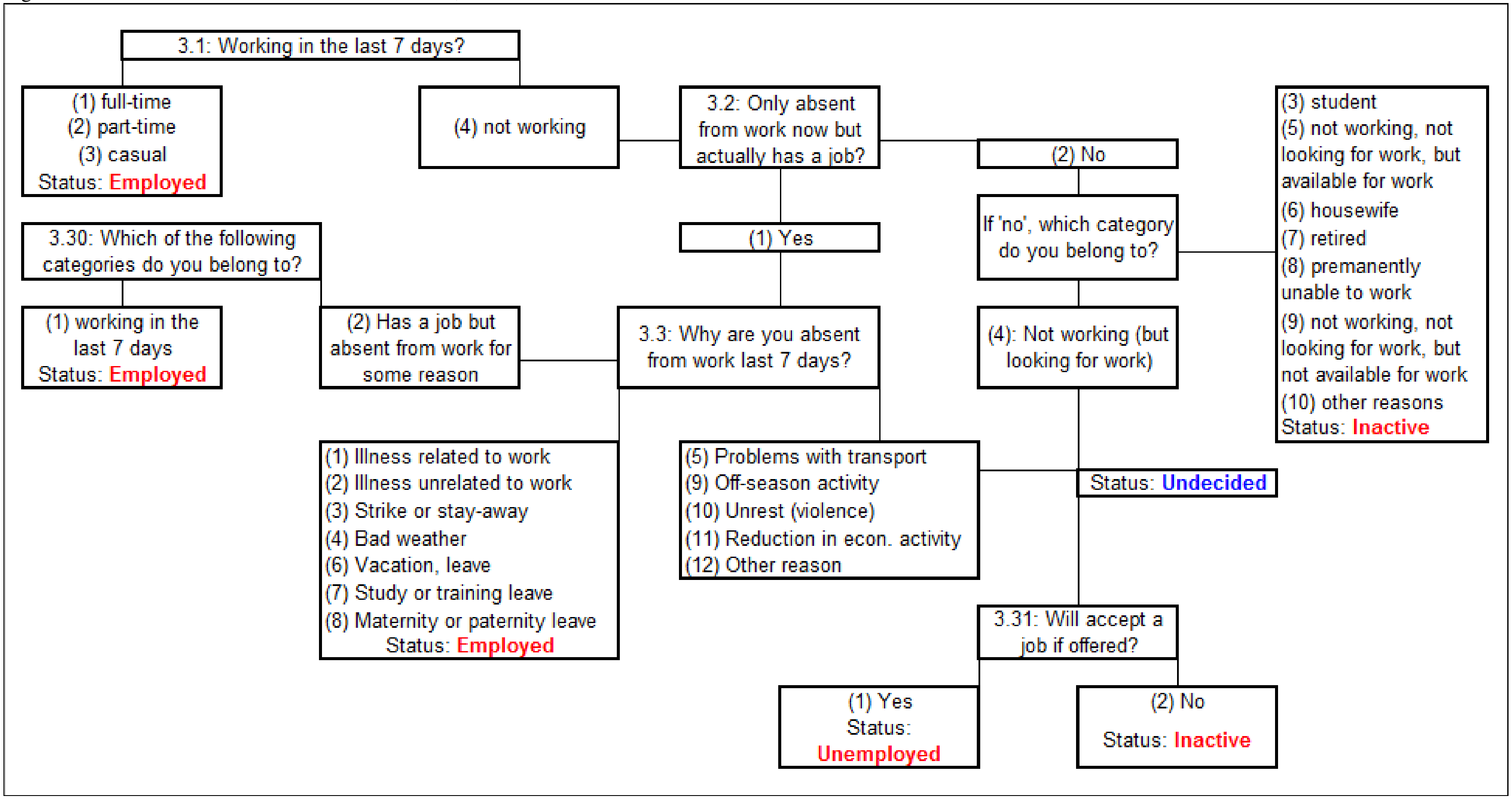
Note: The question number refers to the OHS 1996 questionnaire.

Figure 3.7: Derivation of narrow labour market status, OHS 1996



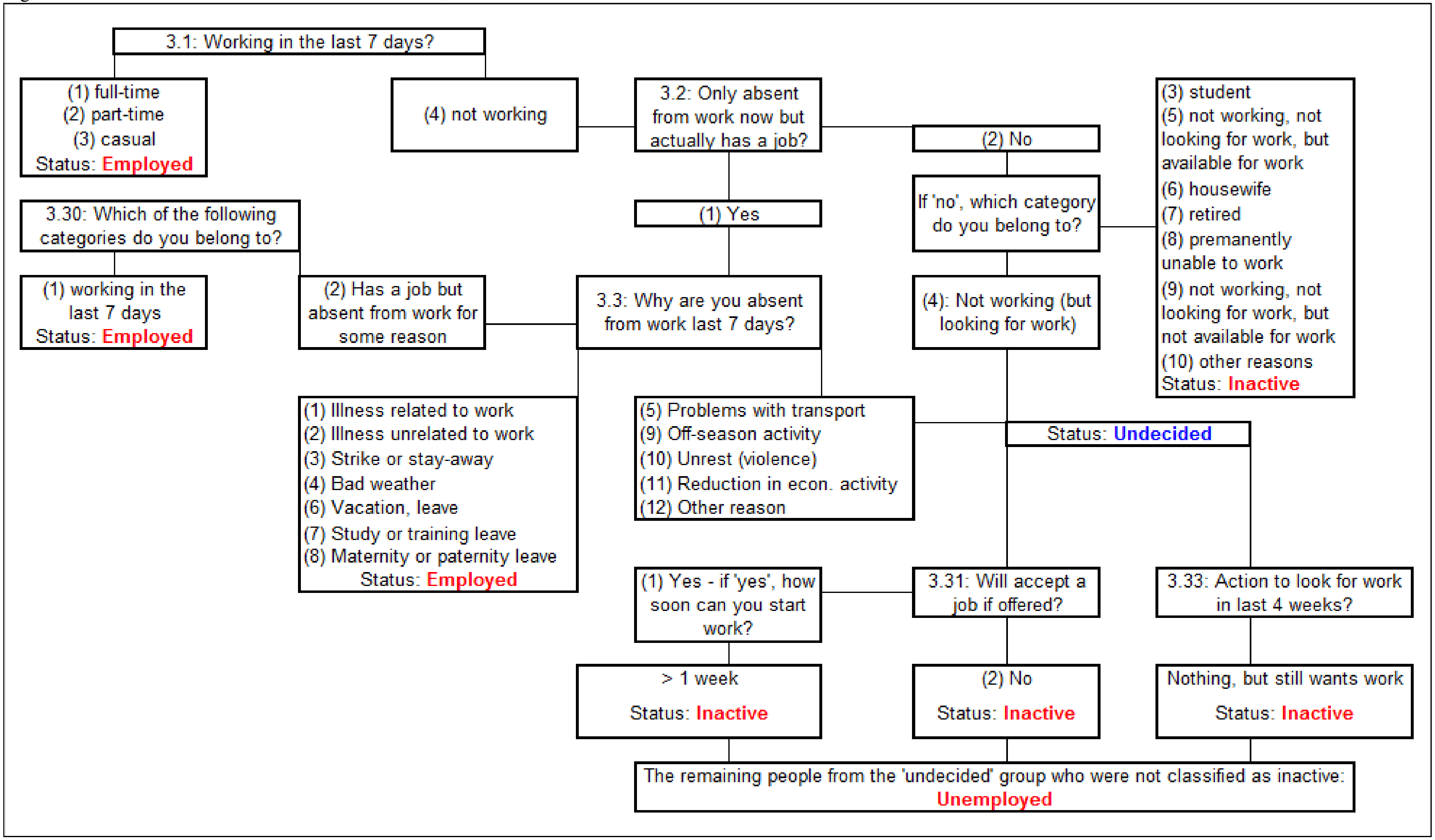
Note: The question number refers to the OHS 1996 questionnaire.

Figure 3.8: Derivation of broad labour market status, OHS 1997-OHS 1998



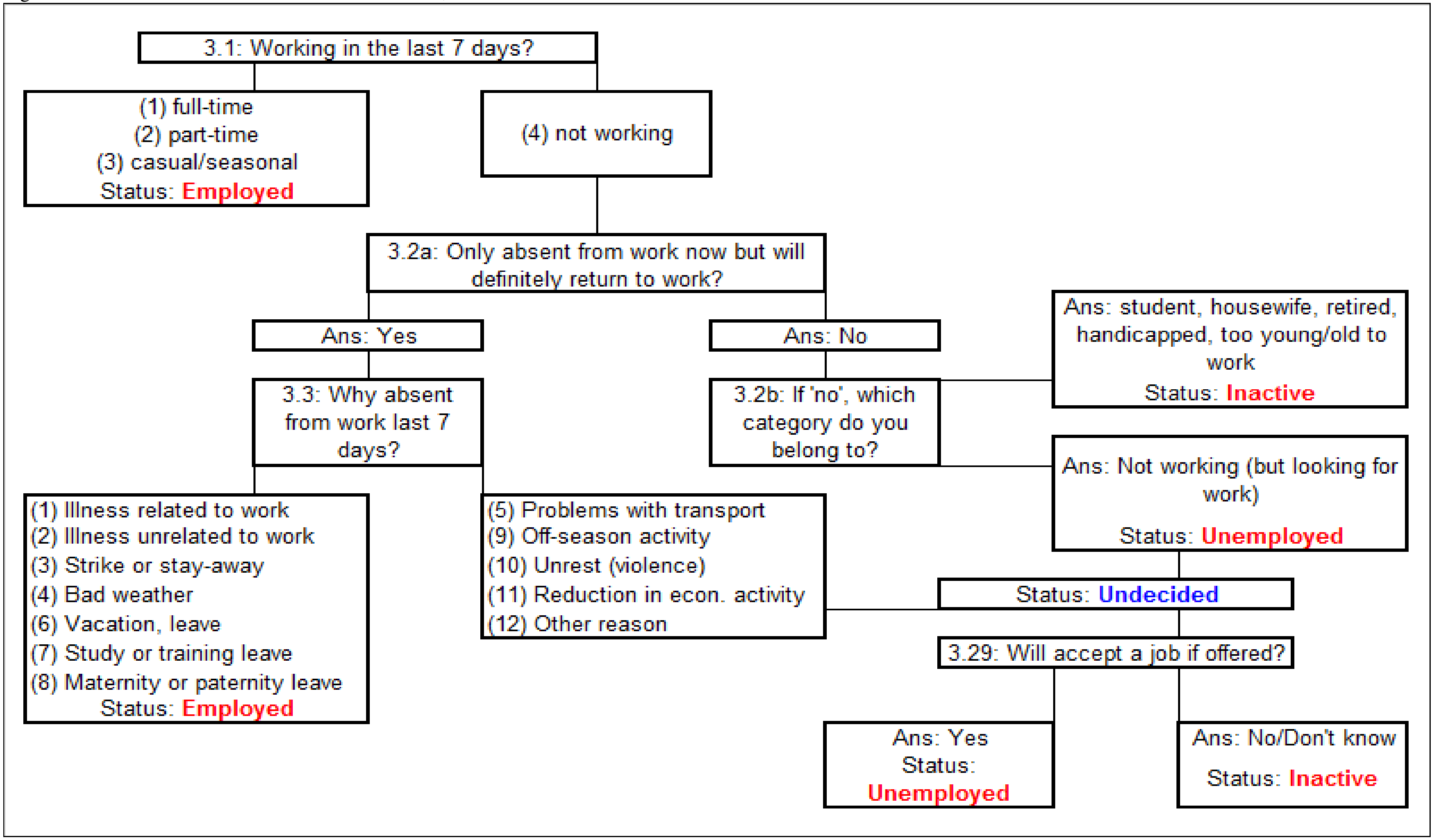
Note: The question number refers to the OHS 1997 questionnaire.

Figure 3.9: Derivation of narrow labour market status, OHS 1997-OHS 1998



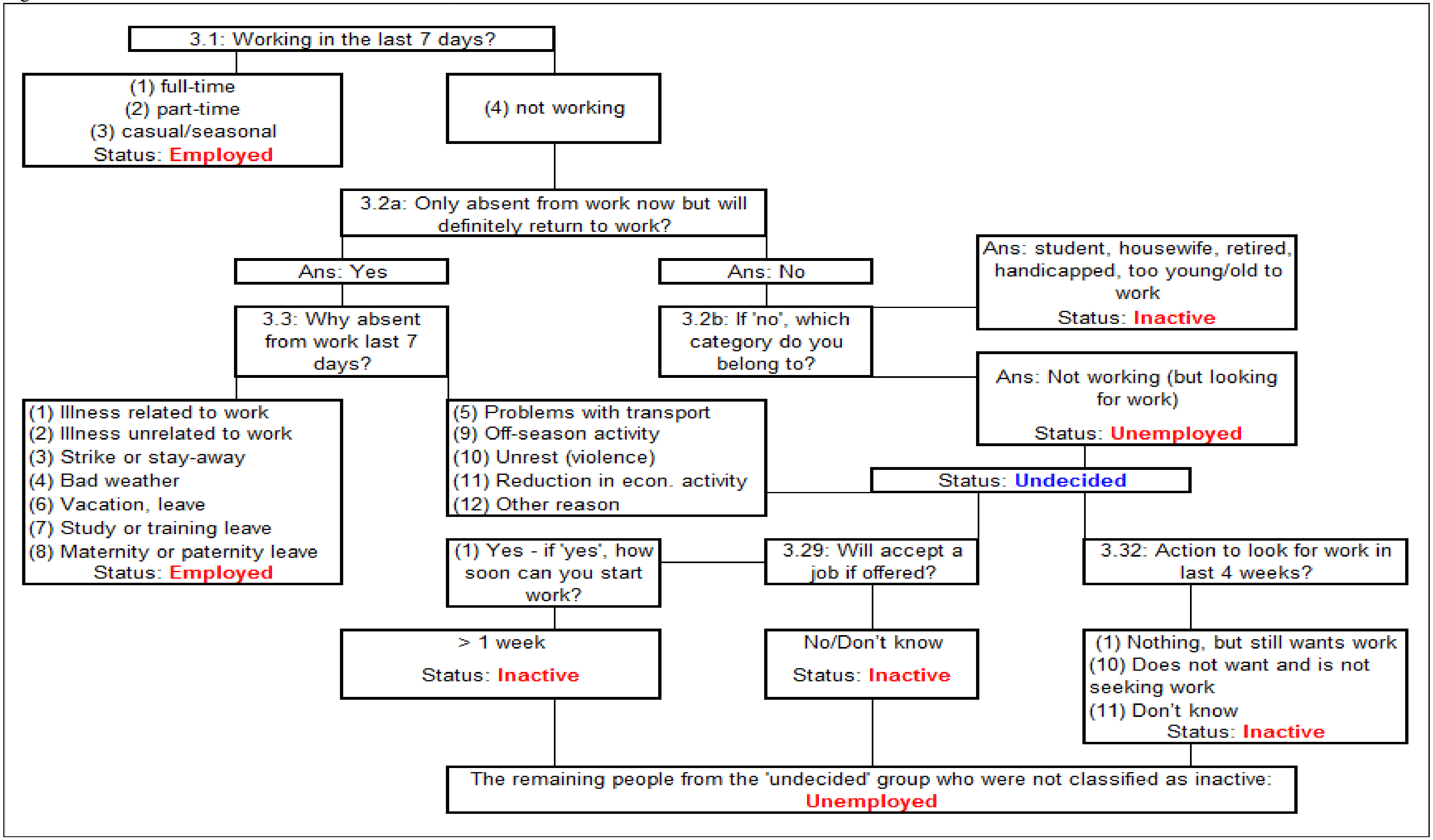
Note: The question number refers to the OHS 1997 questionnaire.

Figure 3.10: Derivation of broad labour market status, OHS 1999



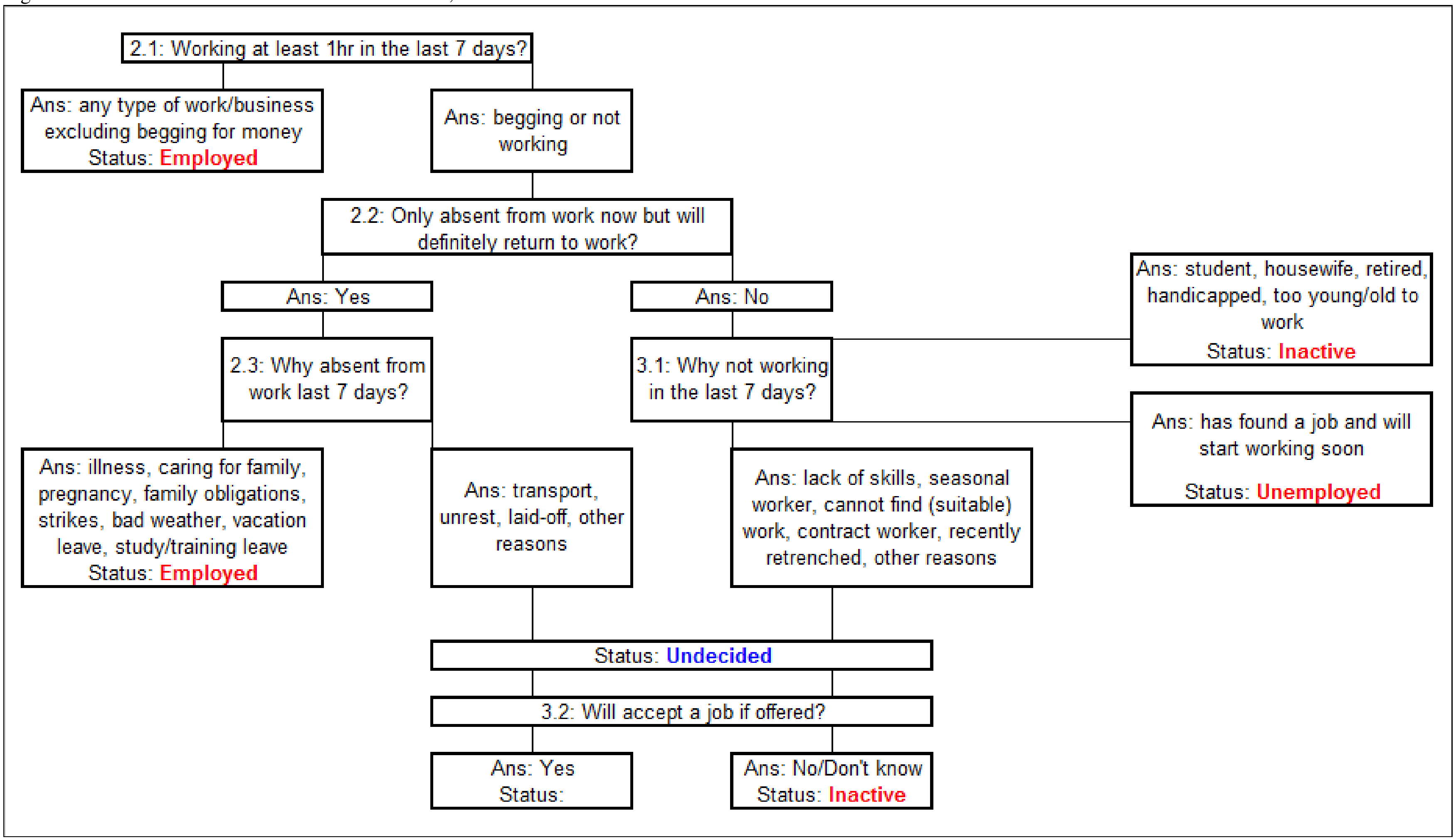
Note: The question number refers to the OHS 1999 questionnaire.

Figure 3.11: Derivation of narrow labour market status, OHS 1999



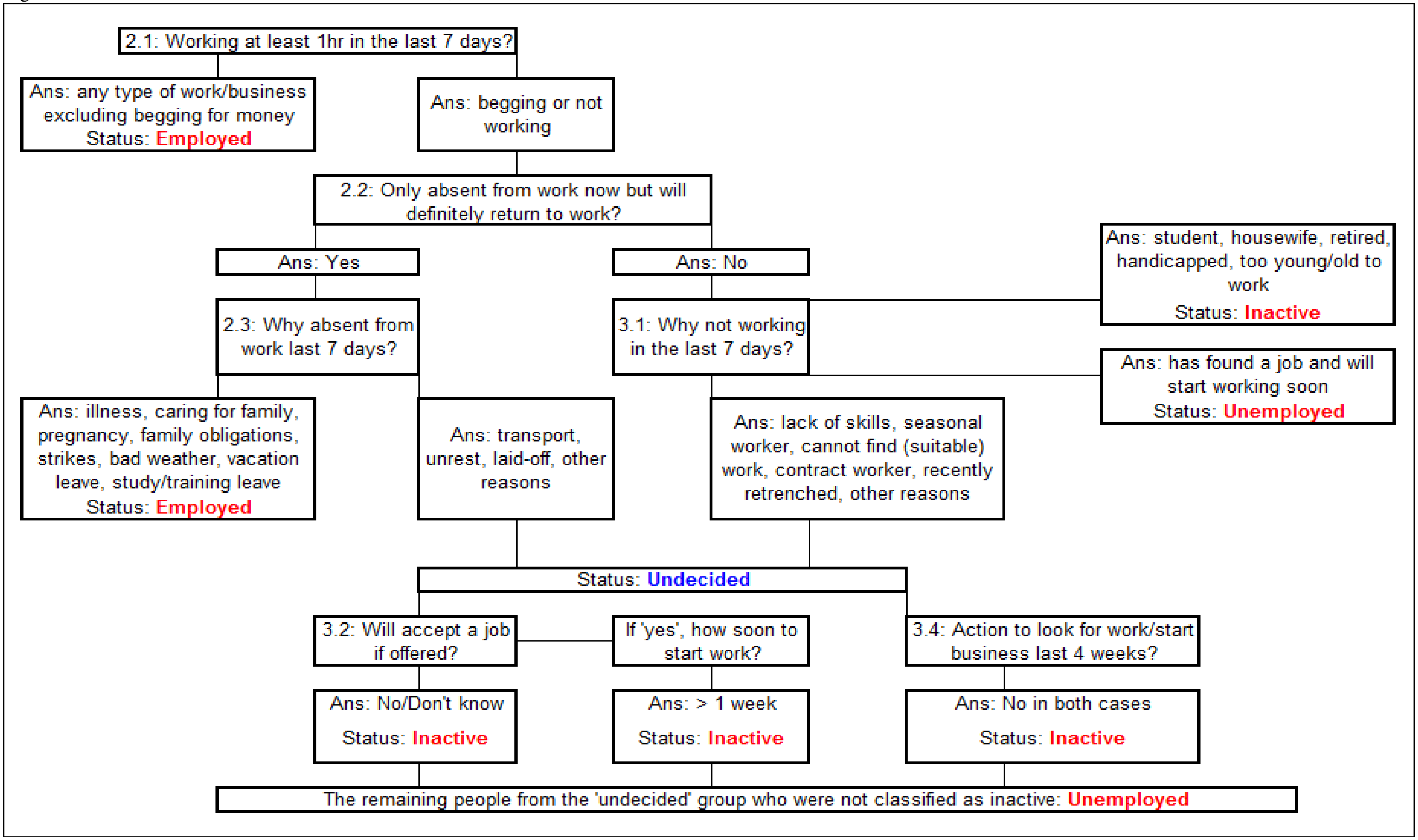
Note: The question number refers to the OHS 1999 questionnaire.

Figure 3.12: Derivation of broad labour market status, LFS 2000a



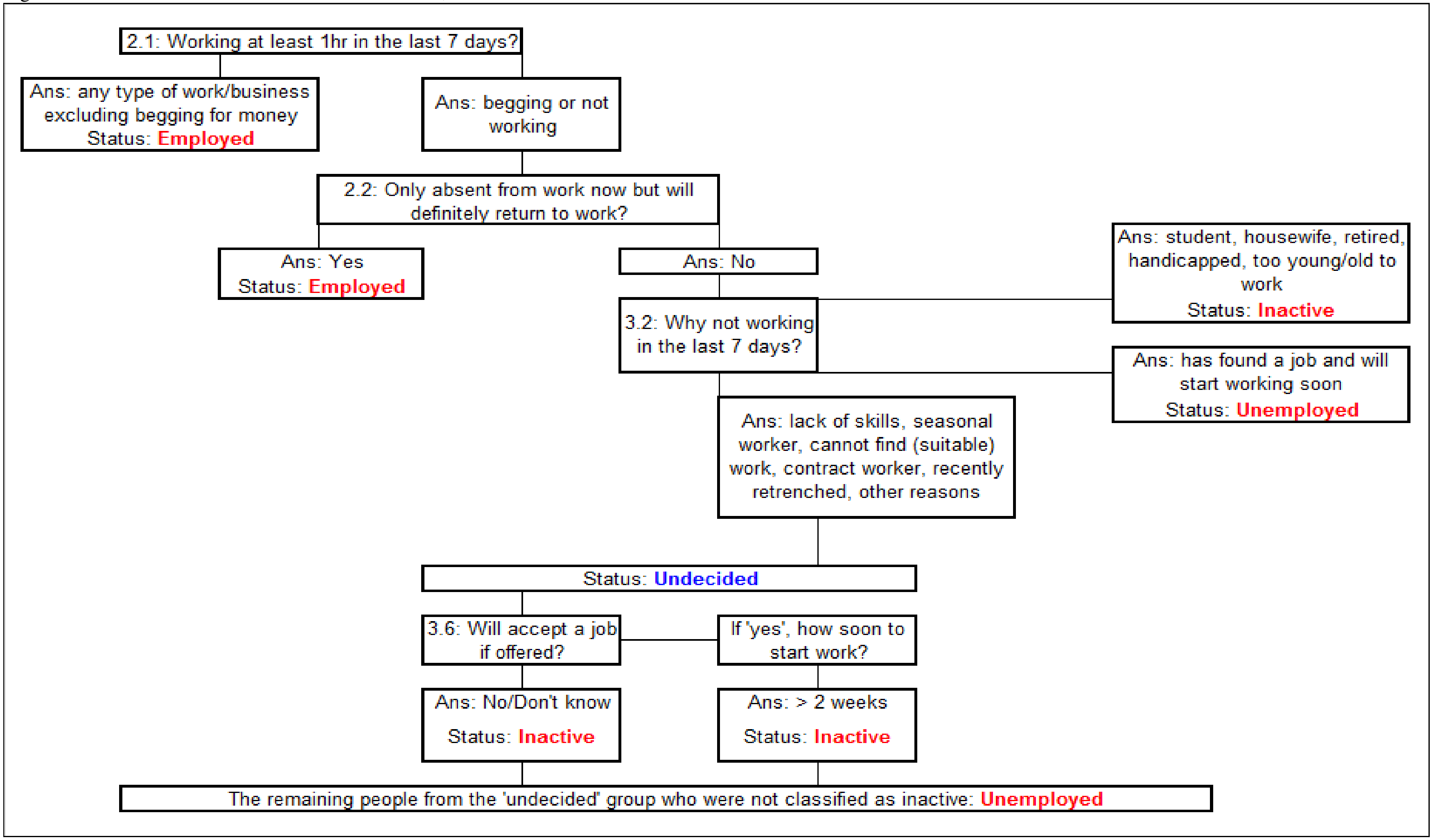
Note: The question number refers to the LFS 2000a questionnaire.

Figure 3.13: Derivation of narrow labour market status, LFS 2000a



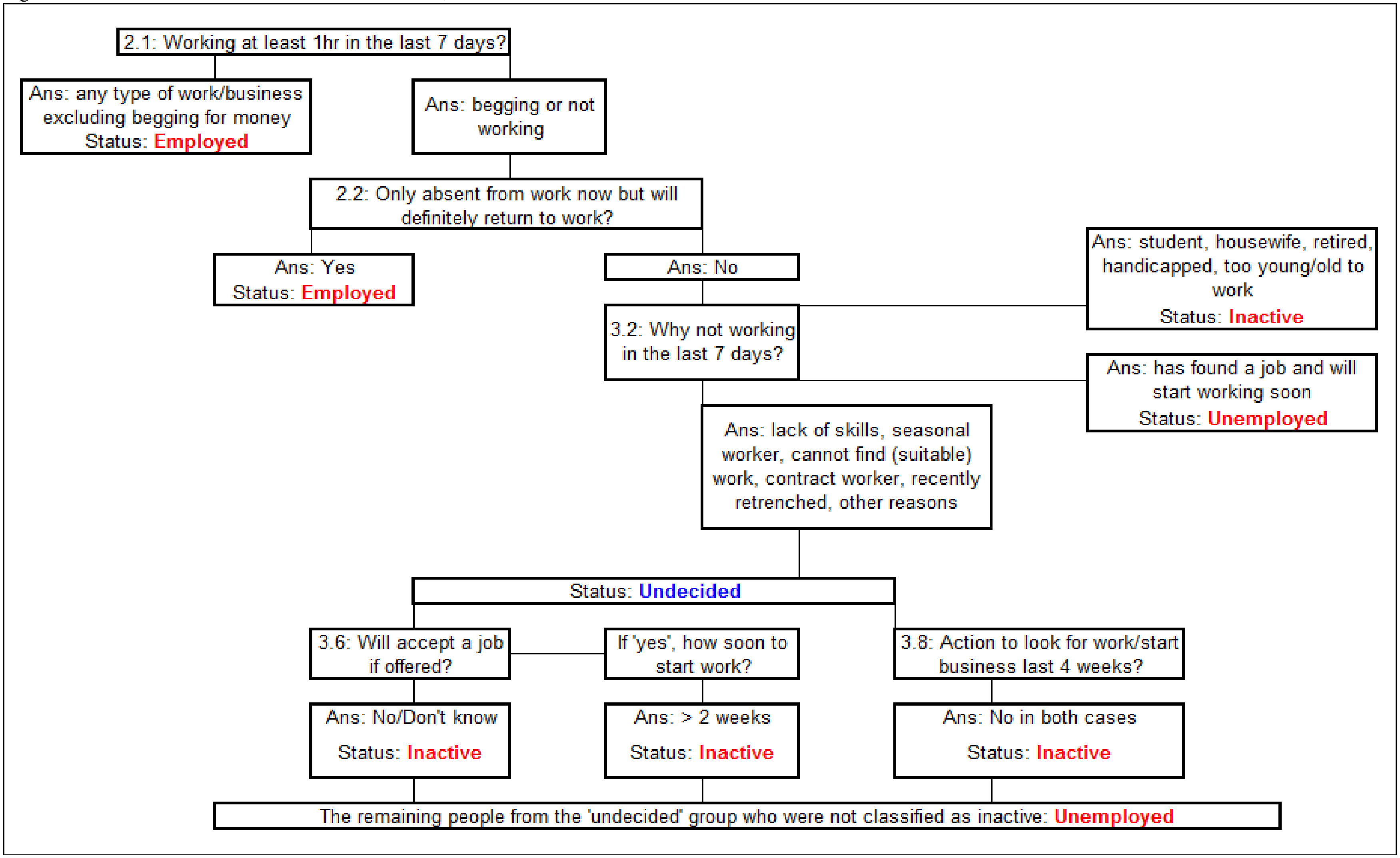
Note: The question number refers to the LFS 2000a questionnaire.

Figure 3.14: Derivation of broad labour market status, LFS 2000b-LFS 2007b



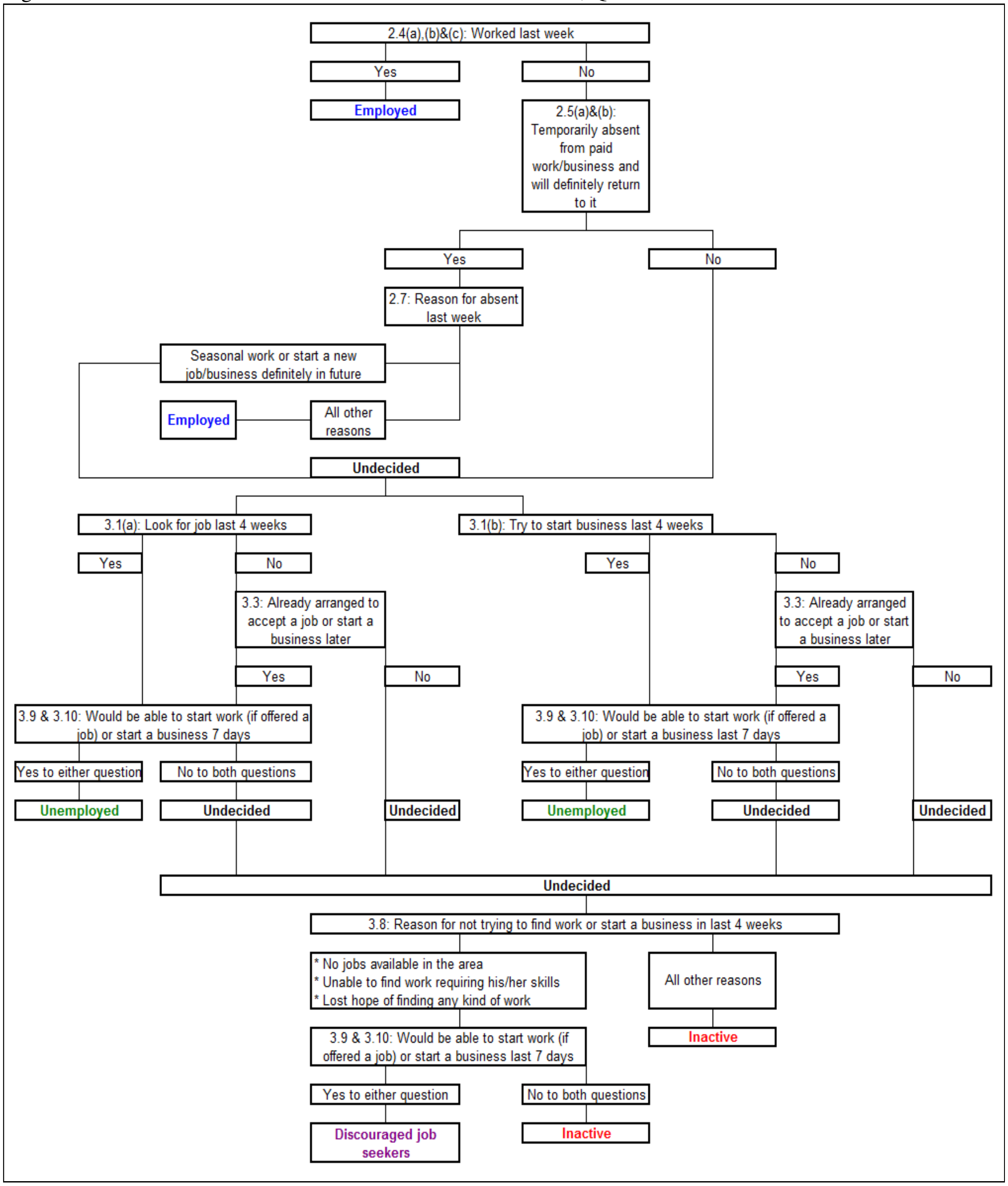
Note: The question number refers to the LFS 2007b questionnaire.

Figure 3.15: Derivation of narrow labour market status, LFS 2000b-LFS 2007b



Note: The question number refers to the LFS 2007b questionnaire.

Figure 3.16: Derivation of narrow and broad labour market status, QLFSs



Note: The question number refers to the QLFS 2008Q1 questionnaire.

With regard to the derivation of employed, in OHS 1996, people reporting as working full-time or part-time during the last seven days were immediately defined as employed (Table 3.6 and Figures 3.6-3.7), while the 1997 and 1998 surveys added a third alternative ‘casual work’ to the questionnaire (Table 3.6 and Figures 3.8-3.9), and in 1999, ‘seasonal worker’ was also added as an option (Table 3.6 and Figures 3.10-3.11). However, the question did not clearly indicate the meaning of the key words ‘full-time’/ ‘part-time’/‘casual work’/‘seasonal worker’. Hence, some respondents might end up reporting they did not work during the last seven days, and eventually be incorrectly classified as inactive or unemployed. For example, someone who worked only two hours per week in the OHS years might think that was not long enough to be called a part-time worker, not to say a full-time worker. If he/she declared he did not engage in any action to seek work, he/she could have been wrongly classified as unemployed or inactive.

Table 3.6: The answer that must be provided by the respondents before they could be qualified as employed, OHS 1993-QLFS 2009Q4

OHS 1994 Now I am going to ask questions about … activities. What did … do most during the last 7 days? 1: Working
OHS 1995 – OHS 1996 Now I am going to ask questions about ... activities. What did ... do most during the last 7 days? 1: Working full-time 2: Working part-time
OHS 1997 – OHS 1998 During the past 7 days, did (the person) do work for pay, profit, or family gain? 1: Yes, full-time 2: Yes, part-time 3: Yes, casual
OHS 1999 During the past 7 days, did (the person) do work for pay, profit, or family gain? 1: Yes, full-time 2: Yes, part-time 3: Yes, casual/seasonal
LFSs In the last seven days, did …… do any of the following activities, even for only one hour? 1: Run or do any kind of business, big or small for himself/herself? 2: Do any work for a wage, salary, commission or any payment in kind? 3: Do any work as a domestic worker for a wage, salary, or any payment in kind? 4: Help unpaid in a family business of any kind? 5: Do any work on his/her own or the family's plot, farm, food garden, cattle post or kraal or help in growing farm produce or in looking after animals for the household? 6: Do any construction or major repair work on his/her own home, plot, cattle post or business or those of the family? 7: Catch any fish, prawns, shells, wild animals or other food for sale or family food?
QLFSs In the last week, 1: Did you work for a wage, salary, commission or any payment in kind (including paid domestic work), even if it was for only one hour? *Examples: A regular job, contract, casual or piece work for pay, work in exchange for food or housing, paid domestic work.* 2: Did you run or do any kind of business, big or small, for yourself or with one or more partners, even if it was for only one hour? *Examples: Commercial farming, selling things, making things for sale, construction, repairing things, guarding cards, brewing beer, collecting wood or water for sale, hairdressing, crèche businesses, taxi or other transport business, having a legal or medical practice, performing in public, having a public phone shop, etc.* 3: Did you help without being paid in any kind of business run by your household, even if it was for only one hour? *Examples: Commercial farming, help to sell things, make things for sale or exchange, doing the accounts, cleaning up for the business, etc.*

From LFS 2000a onwards, the respondents must have worked at least one hour in the last week before they were defined as employed. This approach was consistent with the international definition of employment adopted by the 13th International Conference of Labour Statisticians (ICLS) of the International Labour Organization (ILO) (Haussmans 2007). Furthermore, the respondents were asked if they involved in any one of numerous mostly low-income activities such as ‘guarding cars’ or ‘making things for sale’ (See Table 3.6 as well as Figures 3.12-3.16). It is therefore clear that increased effort was made to capture informal, self-employed and low-income employment. An implication of this improved effort may well be the upward trend in the OHSs as well as the abrupt increase between OHS 1999 and LFS 2000a in the number of employed. Chapter 4 will return to this issue when examining the employment levels and trends since the transition.

In all surveys under study, people who declared they did not work in the last seven days, but also claimed that they were only absent from work at the time of the survey but will definitely return to work were also defined as employed. However, these people only account for a negligible proportion (below 1% in all surveys) of all employed.

From OHS 1996 to LFS 2007b, for those who were not defined as employed but claimed that they were looking for work, they were distinguished as narrow unemployed if they meet the following three requirements (the same approach was adopted internationally by ILO (Haussmans 2007)):

- They will accept a job if being offered one;
- Assuming the job offer is accepted, they could start working within one week. This became more lenient in LFS 2000b-LFS 2007b, as respondents claiming they could start working within two weeks were accepted;
- They took some action to look for work in the last four weeks, such as enquiring at workplaces, placing or answering advertisements, registered at employment agency, etc.

Also, since LFS 2000a, for those people who did not work in the last seven days but claimed they have found a job and will start working soon, they were immediately defined as unemployed.

As far as the derivation of the broad unemployed is concerned, in OHS 1996-LFS 2000a, only the first requirement as discussed above must be met before someone who was not defined as employed but claimed he/she was looking for work would be classified as unemployed.

However, in LFS 2000b-LFS 2007b, the person must meet the first two requirements (i.e., willing to accept a job if being offered one, and being able to start working within two weeks) before he/was would be distinguished as unemployed.

In March 2005, consultants from the International Monetary Fund (IMF) were appointed to evaluate all aspects of the Labour Force Survey and this led to the revision of the labour market status derivation methodology with the launch of the QLFS (Stats SA 2008c). In the QLFSs, the employed were identified in very similarly as in the LFSs (See Table 3.6). The narrow unemployed was also distinguished in a comparable way as in the LFSs in that the respondents (who were not classified as employed) must meet the following three requirements:

- They have already arranged to accept a job or start a business later (the question on starting a business was not asked in the OHSs and LFSs);
- Assuming they have arranged to accept a job or start a business, they could start working or start the business within one week (the question on being able to start the business within one week was not asked in the OHSs and LFSs);
- They took some action to look for work or start a business in the last four weeks.

However, the broad unemployed in the QLFS was identified very differently than in the OHSs and LFSs. In fact, they are defined more strictly in these surveys, as the respondents' answers to the question 'What was the main reason why you did not try to find work or start a business in the last four weeks?' is involved. For the remaining respondents who were not classified as either employed or narrow unemployed, if their reasons for not trying to find work or start a business in the last four weeks were 'no jobs available in the area', 'unable to find work requiring his/her skills' or 'lost hope of finding any kind of work', and they claimed that they could start working (if being offered a suitable job) or start a business within one week, they would be classified as discouraged workseekers, while the remaining people are defined as inactive. Both the narrow unemployed and the discouraged workseekers were in turn defined as broad unemployed. Hence, it is expected that the number of broad unemployed would decrease suddenly during the changeover from LFS to QLFS.

To conclude, the changes in the labour market status derivation could lead to some abrupt changes in the labour market aggregates since the advent of democracy. Chapter 4 will examine whether the levels and trends of the labour market aggregates would be significantly different if a consistently labour market status derivation methodology is applied across the surveys. For example, the LFS 2000b-LFS2007b methodology is applied to all OHSs, or the QLFS methodology is applied to all OHSs and LFSs.

3.12 Other factors

Other issues such as the length of the questionnaire (which is influenced by the level of disaggregation of income and expenditure items), whether the labour market status, income and expenditure questions were asked at the beginning or rather the end of the questionnaire (as the interviewees might feel tired towards the end of the interview), quality of training received by the interviewers prior to the start of the interviews, their experience and efforts devoted to capture information during the interviews could also affect the comparability of the various datasets.

Table 3.7 shows that the length of the questionnaire is relatively longer (at least 50 pages) in the IESs, GHSs, PSLSD, NIDS, AMPSs and OHS 1995-1999. There are many more questions relating to income and expenditure in the IESs, due to the greater level of disaggregation of items when capturing information on total household income and expenditure. In addition, as expected, there are more precise questions asked to capture labour market information in the OHSs, LFSs and QLFSs. Hence, it is possible that the interviewee or even interviewer fatigue could happen if the questionnaire is too long[54].

With regard to the surveys that captured the income or expenditure information by only asking one question to the respondents to declare the 'one-shot' amount in either actual amount or interval terms, it is found that this question was asked almost at the end of the questionnaire in the OHSs, LFSs and AMPSs, and as indicated in Table 3.7, the questionnaire is very long in these surveys. It is not known whether the respondents might have felt extremely tired towards the end of the interview that they might not report accurate answers on their income or expenditure.

Chapters 5 and 6 will come back to these issues when dealing with the derivation of total income and expenditure, as well as poverty and inequality estimates.

[54] As discussed in Section 2.3, IES 2005/2006 addressed this issue by dividing the questionnaire into five parts, with the interviewer visiting the responding household five times and asking questions from each part in each visit.

Table 3.7: More information on the income, expenditure, and labour market status questions in each survey

Survey	Year	Length of the questionnaire	Page number where income / expenditure questions were asked	Total number of pages of income / expenditure / labour market status questions
			Income / Expenditure question(s)	
Census / CS	1996	10 pages	P.6	1 page (1 question only)
	2001	6 pages	P.3	1 page (1 question only)
	2007	21 pages	P.18	1 page (1 question only)
IES	1995	30 pages	Expenditure: P.6-26 Income: P.27-28	Expenditure: 21 pages Income: 2 pages
	2000	67 pages	Expenditure: P.8-47 Income: P.51-56	Expenditure: 40 pages Income: 6 pages
	2005/2006	Interview #1: 12 pages	N/A[#]	None
		Interview #2: 13 pages	Expenditure: P.5-13	Expenditure: 9 pages
		Interview #3: 16 pages	Expenditure: P.1-16	Expenditure: 16 pages
		Interview #4: 12 pages	Expenditure: P.1-12	Expenditure: 12 pages
		Interview #5: 14 pages	Expenditure: P.1-4 Income: P.5-14	Expenditure: 4 pages Income: 10 pages
OHS	1994-1999	1994: 30 pages 1995-1999: 54-67 pages	1994-1995: N/A[#] 1996-1999: Expenditure: Almost at the end of questionnaire	1994-1995: N/A 1996-1999: 1 page (1 question only)
LFS	2001-2007	25-35 pages	2001-2004: Expenditure: Almost at the end of questionnaire 2005-2007: N/A[#]	2001-2004: 1 page (1 question only) 2005-2007: None
QLFS	2008-2009	13 pages	2008-2009: N/A[#]	2008-2009: None
GHS	2002-2009	About 50 pages	2002-2009: Expenditure: Almost at the end of questionnaire	2002-2009: 1 page (1 question only)
PSLSD	1993	57 pages	Expenditure: P.10-17 Income: P.32, 36, 39, 48, 50	Expenditure: 8 pages Income: 5 pages
NIDS	2008	Household: 27 pages Adult: 32 pages	Expenditure (Single estimate): P.9, household questionnaire Income (Single estimate): P.11, household questionnaire Expenditure (Aggregate): P.12-16, household questionnaire Income (Aggregate): P.11 & 14, adult questionnaire	Expenditure (Single estimate): 1 page Income (Single estimate): 1 page Expenditure (Aggregate): 5 pages Income (Aggregate): 2 pages
AMPS	1993-2009	About 75 pages	Income: +-P.60	Income: 1 page (1 question only)
			Labour market status questions	
OHS	1994-1999	1994: 30 pages 1995-1999: 54-67 pages	From about P.20 onwards	15-20 pages
LFS	2001-2007	25-35 pages	From about P.12 onwards	15-20 pages
QLFS	2008-2009	13 pages	P.5-13	9 pages

[#] None of the questions are related to income or expenditure.

3.13 Chapter summary

This chapter first discussed the arguments for and against using income and expenditure (consumption) for poverty and inequality analyses. Although the general conclusion is that expenditure is the preferred measure to be used in developing countries, further investigation shows that this might not be the case. Secondly, the advantages and disadvantages of using the traditional recall approach and the diary approach to capture the income and expenditure information were discussed, and it seems durable expenditure would always be captured with some flaws, regardless of which approach is adopted.

The issue of whether the income and expenditure should be captured in actual amounts or in bands / intervals / categories was investigated, as each method involves advantages and disadvantages. If the information is collected in actual amounts, the next question that arises is whether the amounts should be captured as a 'one-shot' single estimate or rather the aggregation of amounts from different sources. The pros and cons of each approach were discussed. If the information is collected in intervals instead, three issues come up: The appropriate method to convert the interval data into continuous data for the subsequent poverty and inequality analyses; the impact of the number of bands and width of each band on the poverty and inequality estimates; and how to deal with households with zero or unspecified income or expenditure. It was found that the midpoint-Pareto method was most appropriate to make the interval data continuous. Moreover, there is insufficient research done both domestically and internationally that investigate how the number and width of bands affect the poverty and inequality estimates. Furthermore, the sequential regression multiple imputation (SRMI) approach is used to impute the income (or expenditure) of households reporting zero or unspecified income (or expenditure).

The possible merits and drawbacks of adjusting the survey income (or expenditure) distribution in line with the national accounts income mean, as well as the validation of the survey data against external sources (e.g., income tax revenue data by the National Treasury) to evaluate the reliability of the former data were discussed. Furthermore, since the post-stratification adjustment of the survey weights in the Stats SA survey datasets did not take account of temporal consistency issue, concerns have been raised with regard to using these cross-sectional datasets to investigate the change of labour market, poverty and inequality estimates over time. The cross entropy approach could address the temporal inconsistency problems and the minimum cross entropy (CE) will be adopted in later chapters to re-weight the datasets for further analyses on the aforementioned estimates over time.

Chapter 3 also discussed how the labour market status derivation methodology has changed over the years, with particular attention on the drastic changes to the broad methodology since the introduction of QLFS. How these changes affect the comparability and reliability of labour market aggregates across the surveys will be dealt with in Chapter 4. Finally, the length of the questionnaire, the number of questions to capture the income, expenditure and labour market status information as well as whether these questions were asked at the beginning or rather towards the end of the interview also play a role to affect the accuracy of these information, and subsequently the poverty, inequality and labour market estimates.

Chapter 4 will examine the labour market trends while Chapters 5 and 6 will investigate the poverty and inequality trends since the transition, paying particular attention to how the issues discussed in Chapters 2 and 3 could affect the reliability and comparability of these trends across the surveys.

CHAPTER FOUR: LABOUR MARKET TRENDS IN SOUTH AFRICA SINCE THE TRANSITION

4.1 Introduction

Many recent studies (e.g., Casale & Posel 2002; Bhorat 2004; Burger & Woolard 2005; Oosthuizen 2006; Van der Westhuizen et al. 2006) reviewed the South African labour market 'trends' by comparing the OHS 1995 data with the most recent available OHS or LFS data at the time of writing. However, OHS and LFS are incomparable in many aspects, given changes in the sampling frame, inconsistencies in questionnaire design, coding errors, changes in methodology to capture employment status, outliers in wage earnings data, etc[55]. Furthermore, comparing an OHS with an LFS provides only a snapshot of the South African labour market at two points in time, but does not provide detail on the labour market trends over the period. Hence, this chapter aims to give a more detailed picture of the labour market trends from 1994 to 2009[56], using all the available OHS, LFS and QLFS data, since these are the surveys with the primary aim of capturing labour market status, as mentioned in Chapter 2[57]. Such methodology avoids the problem of a two-snapshot overview, whilst allowing a clearer picture of the trends in the labour market over the period in question.

Section 4.2 reviews the recent South African studies on the labour market trends since the transition, while Section 4.3 focuses on the demographic, geographic and educational attainment characteristics of the labour force (LF) and trends in labour force participation rate (LFPR), considering whether increased 'feminisation of the labour force' as suggested by various recent studies (Casele & Posel 2002; Casale 2004; Burger & Woolard 2005; Oosthuizen 2006; Van der Westhuizen et al. 2006) took place or not. Section 4.4 discusses

[55] Most of these problems were discussed in Chapters 2 and 3.

[56] OHS 1993 will not be analysed, since the sample did not include people from the former TBVC states, as mentioned in Chapter 1.

[57] Figures A.1 – A.4 in Appendix A compare the narrow and broad LFPRs as well as the narrow and broad unemployment rates of different surveys since 1993. PSLSD, which only captured the narrow labour market status, reported lower LFPR and unemployment rate (compared with the OHS rates). Looking at the results from censuses and CS 2007, Census 2001 seriously under-estimated the broad LFPR but over-estimated the broad LFPR and unemployment rates, when compared with the LFS 2001 figures. Furthermore, CS 2007 over-estimated the narrow LFPR and unemployment rates in comparison with the 2001 LFS figures. It was expected that PSLSD, censuses and CS 2007 might not have captured the labour market status of respondents well, because the capture of the above information was not the primary aim of these surveys (See Chapter 2).

NIDS, despite adopting the same methodology as in QLFS to derive labour market status, recorded higher LFPRs and unemployment rates in both narrow and broad terms, when compared with the QLFS 2008 figures. Finally, looking at GHSs, which adopted the same labour market status derivation methodology as in LFS 2000b-2007b, the LFPR and unemployment rate figures and trends are very close to those in the LFSs, except that the GHS broad LFPRs are always about 2 percentage points lower than the LFS rates.

employment trends, with specific reference to occupation, industry, skills and formal/informal employment status. Characteristics of the unemployed are looked at in Section 4.5.

Even if the all the OHSs, LFSs and QLFSs are taken into consideration to examine the labour market trends throughout the years, it does not mean this would entirely solve the incomparability issue across the surveys. In fact, the abrupt changes (e.g., the possible sudden decline of the number of broad unemployed between LFS 2007b and QLFS 2008Q1, as discussed in Section 3.11) might still have taken place even after looking at all available labour surveys. Hence, Section 4.6 adopts various approaches to improve the reliability and comparability of labour market aggregates across the surveys. First, the OHS/LFS/QLFS data are compared with other employment data, with particular focus on whether the abrupt trends across some surveys is attributed to the fluctuations of the self-employment and informal sector employment, i.e., whether the comparability would improve if only non-agricultural employees are looked at. Secondly, the labour market trends are re-visited after the consistent cross entropy approach is adopted to re-weight all the datasets. Finally, whether the application of the same labour market status derivation methodology across all the surveys, if possible, could lead to the derivation of more precise and reliable labour market aggregates is investigated. Section 4.7 concludes the chapter.

Other important issues such as the causes of unemployment[58] and the policies which aim to solve the unemployment problem[59] (Mahadea 2003; Kingdon & Knight 2004 & 2007; Edgren 2005; Arora & Ricci 2006; Banerjee, Galiani, Levinsohn and Woolard 2006; Pauw et al. 2006; Centre for Development and Enterprise 2007; Bhorat 2009) fall beyond the scope of the dissertation and will not be discussed.

4.2 Literature review of recent studies on labour market trends

In general, there have been three types of studies on the South African labour market trends since the advent of democracy:

- Studies that compared OHS 1995 with the latest available OHS or LFS at the time of

[58] These causes include the following: labour force participation increased too rapidly since the transition; mismatch of skills supplied by the labour force participants and the skills demanded by employers; wage rigidity due to trade union pressure and collective bargaining; employment inflexibility due to the labour market legislations like Affirmative Action and Employment Equity Act; barriers of entry to informal sector; lack of employment prospects of youth graduates (due to reasons such as their wrong field of study, relatively poor quality of education received by many graduates from historically black tertiary institutions, lack of soft or social skills), and high reservation wage.

[59] For example, promoting medium and small-scale enterprise, skills development programs, etc.

writing to derive labour market 'trends' on measures like LF and their characteristics, LFPRs, employment and the characteristics of the employed and unemployed, as well as unemployment rates. However, strictly speaking, this approach could only provide a snapshot of the labour market at two points in time and not trends.

- Studies that used most of or all available datasets at the time of writing to derive labour market trends over the period under study, the variables of interests being the same as those in the first type of study.
- Studies that used most of or all available datasets at the time of writing, but focused on investigating the employment gap, wage trends and wage gap, to determine if labour market discrimination in terms of employment discrimination and wage discrimination remained significant since the transition. Econometric techniques like two-step Heckprobit and Heckman earnings regressions, as well as Oaxaca and Blinder decomposition have been used.

4.2.1 Review of studies that compared two surveys to derive labour market 'trends'

The studies by Poswell (2002), Bhorat (2004, 2005, 2006 & 2009), Bhorat and Oosthuizen (2005), Burger and Woolard (2005), Oosthuizen (2006), Dias and Posel (2006) are more general, comparing OHS 1995 with the latest OHS or LFS at the time of writing to look at the labour market 'trends' since the transition. In contrast, studies by Casale and Posel (2002), Casale (2004), Van der Westhuizen et al. (2006), Dias and Posel (2006) and Pauw et al. (2006), despite adopting the same approach, had clearer research objectives, with the first three studies focusing on whether feminisation of labour market took place since the advent of democracy, Dias and Posel examining the relationship between education and unemployment, and Pauw et al. focusing on graduate unemployment problem.

Looking at the studies that investigated the labour market 'trends' in general, Poswell (2002) compared OHS 1995 with OHS 1999, mainly focusing on the characteristics of the employed. She found that employment increased for all race and gender groups over the period; the increase was most rapid in the white and female population. However, the growth of employment was inadequate to absorb the large increase in the number of LF. Hence, this resulted in increasing rates of unemployment across all races and gender. Poswell argued that the key challenge for the economy was to match the increase in demand for highly-skilled workers with an adequate supply, but the country did not possess the necessary expertise. The situation was exacerbated by the emigration of highly-skilled professionals and the increasing impact of HIV/AIDS.

Bhorat, in his 2004 and 2006 articles, compared OHS 1995 with LFS2002a to identify the key trends in the labour market in an attempt to understand the factors that might be driving the performance of this factor market. The focus was on the narrow definition of the LF. He investigated whether the so-called jobless growth phenomenon as defined by Altman (2003: 12) took place in this period. He found that while South Africa did not have jobless growth under the first definition (i.e., real GDP increase in conjunction with a stagnant or decline in the absolute employment level), but the employment was clearly insufficient relative to the growth in the LF, resulting in jobless growth under the second definition (i.e., real GDP increase accompanied by an increasing unemployment rate), since the narrow unemployment rate showed an upward trend. Bhorat also found that the nature of employment growth was biased towards highly-skilled and semi-skilled workers, with unskilled workers more likely to be retrenched over the period under study. In addition, the unemployment rate was greater among blacks, females, and those without Matric. Most of the unemployed came from households with zero or only one employed member. Bhorat (2006) also analysed the LFS 2002b data by running an OLS regression on the household narrow unemployment rate. The results showed that this rate was higher if household expenditure was low, the household stayed in a rural area, the household was headed by a non-black female, recipients of the old-age pension and/or child support grant were present, there were fewer trade union members, and if the households experienced hunger[60] more frequently.

In his study to compare OHS 1995 with LFS 2003b, focusing on the narrow definition once again, Bhorat (2005) found that the relative position of blacks, females and the unskilled had declined as these groups were increasingly more likely to be casually employed. In addition, one of the most prominent responses to the labour legislative regime was to increase the portion of atypically employed. Finally, in the most recent study by Bhorat (2009), which compared OHS 1995 with LFS 2005b, two-step Heckman probit regressions[61] on employment probability under the broad definition were run. The results showed that whites, females,

[60] In LFS 2002b (Question 7.22 of Section 7), the household heads were asked if the households had problems satisfying their food needs in the past 12 months, and they could choose from the following five categories: "1: Never", "2: Seldom", "3: Sometimes", "4: Often" and "5: Always".

[61] Since not everyone in the working-age population joined the LF and eventually found employment, the results of a probit regression on employment likelihood of the working-age population would be biased due to sample selection. The most common technique applied to address this problem is a two-step Heckprobit model. The first step is a probit analysis to identify the factors determining whether someone in the working-age population would join the LF or not. The equation allows the estimation of the inverse Mills ratio (i.e., lambda), which is in turn included in the employment probit (i.e., the second step), making the latter regression conditional on labour force participation. If the inverse Mills ratio variable is statistically significant in this probit, it indicates that the labour force indeed differ from their counterparts who decided not to participate in the labour force, and the two-step Heckman approach is necessary.

people aged 25-55 years, with at least Matric, those staying in rural areas and Gauteng were more likely to be employed in 1995, but in 2005, it was whites, males, elderly aged 45 years or above, those with post-Matric qualifications, and those staying in rural areas in Western Cape or Gauteng more likely to be employed. The 2005 results highlighted the increasing youth unemployment problem. To conclude, all four studies by Bhorat implied that the situation of females and blacks deteriorated since the transition, as they became more likely to be unemployed.

Burger and Woolard (2005) investigated the state of the labour market after the first decade of democracy by comparing OHS 1995 with LFS 2002a. The primarily focus was on the characteristics of the unemployed under the broad definition. They found that unemployment continued to rise as job creation did not match the growing labour supply. In addition, the unemployed were predominantly blacks and coloureds, and more than half of the unemployed were under the age of 35 years. Women and rural workers were also over-represented amongst the unemployed. Educational attainment appeared to be an important factor to determine labour market outcomes, since having completed secondary or some form of post-Matric education substantially increased the probability of being employed. However, even the unemployment rates of people with Matric or post-Matric qualifications increased between the two surveys. Furthermore, Burger and Woolard categorized the unemployed into the following three groups:

- The youngest group with complete secondary or post-secondary education. Demand-side policies to stimulate higher economic growth and supply-side policies such as vocational training could help improving their chance of employment.
- Older group (at least 35 years), African females with very low education and no employment experience. It was argued that these people were most likely to be unemployable.
- Older individuals with incomplete secondary education and some forms of labour market experience. Some form of skills upgrading was required before these people could be absorbed into the labour market again.

Finally, Burger and Woolard found that employment growth (1.7 million) across the period was substantial, but many of these employment opportunities were concentrated in the informal sector, where working conditions were inferior to those in the formal sector.

Bhorat and Oosthuizen (2005) compared OHS 1995 with LFS 2002b, analysing the characteristics of the LF, employed and unemployed under the broad definition. The two snapshots of the LF showed that the growth in the LF could be ascribed to predominantly black new job-seekers who were increasingly female, residing in urban areas with some level of secondary education and under the age of 35 years. In addition, the broad LFPR increased relatively faster across the two surveys for the following groups: blacks, females, people from younger age cohorts, and people from Limpopo. Secondly, although employment increased across the period, it was outstripped by the more rapidly expanding LF. Real GDP and employment growth were complemented by the increase of unemployment. Moreover, the characteristics of the employed were analysed, and most of the increase of employment accrued to blacks, females, those aged 35-54 years, and people with at least Matric. Thirdly, unemployment rates by demographic characteristics were investigated. Female rates were always higher in all race groups. In addition, the unemployment rate of people with tertiary education showed an increase, especially amongst blacks. Furthermore, unemployed individuals were increasingly marginalised in households with no wage or salary earnings, which in turn increased the demand placed on elderly household members' state old-age pensions and other grants. Finally, the paper concluded by running simple probit regressions to determine the likelihood of being unemployed in both surveys.

Oosthuizen undertook a similar study in 2006 by comparing OHS 1995 with LFS 2004b, also focusing on the broad definition, and had very similar findings as Bhorat and Oosthuizen (2005). Furthermore, simple probit regressions and two-step Heckprobit regressions were run on labour force participation and unemployment respectively. The results showed that people from the younger age cohorts, blacks, males, those with higher educational attainment and from provinces other than Eastern Cape and Limpopo were more likely to participate, while people aged 15-24 years, blacks, females, those without Matric and those coming from Limpopo were less likely to be employed.

With regard to studies that focused on specific groups of people, Casale and Posel (2002), Casale (2004) and Van der Westhuizen et al. (2006) paid particular attention to the female population. First, Casale and Posel (2002) compared OHS 1995 with OHS 1999 by first presenting some brief statistics on the female LFPRs and unemployment rates under both narrow and broad definitions as well as employment type (formal/informal sector) of the employed. They found that the feminisation of LF took place, but this was complemented by the feminisation of generally insecure forms of employment in the informal sector, as well as increasing female unemployment rates. The second part of the paper focused on the causes of more females looking for work. It was found that the two causes were the increased educational attainment of women as well as the fall in access to male income due to reasons like male unemployment, the deepening of the HIV/AIDS epidemic, an increase in female headship, and an increase in the proportion of females either living with a partner, divorced or separated, or never married.

The study by Casale (2004) investigated what the rise in female employment bought them in terms of access to different types of employment between OHS 1995 and LFS 2001b. The results showed that while white women were the main beneficiaries, most of the growth of employment among the black women took place in unskilled informal self-employment and domestic work that offered extremely low and decreasing returns as well as little security. Furthermore, the feminisation of the LF was still associated with a deterioration of their disadvantaged position in the labour market relative to men between the two surveys.

Van der Westhuizen et al. (2006) investigated trends in the status of women in the labour market between OHS 1995 and LFS 2005b, focusing on the broad definition. They found that the broad female LFPRs in all four race groups, although still below the male LFPRs, showed a more rapid increase between the two surveys, with the increase in LF being driven specifically by greater numbers of black women entering the LF. Moreover, although the female employment as well as the female share of the employed increased, the bulk of the increase in female employment took place in elementary occupations. Finally, female unemployment rates increased for all covariates, but black women and young women in particular struggled to find employment. They argued that the increase in the number of jobs accruing to women was not nearly enough to absorb all the additional female entrants to the labour market. The paper concluded with multivariate regressions on female labour force participation likelihood, employment likelihood and earnings of the employed. The Heckman approach was adopted in the last two regressions so as to deal with the sampling selection bias issue (see Footnote 61). It was found that females continued to suffer discrimination in the labour market, as characterised by lower quality employment and lower remuneration, especially for black females.

Pauw et al. (2006) investigated the graduate unemployment problem by comparing OHS 1995 with LFS 2005b, focusing on broad unemployed. The results showed that the LF and the employed became more educated on average, but unemployment rates increased across all educational attainment categories, confirming the findings by Bhorat and Oosthuizen (2005) and Burger and Woolard (2005) as discussed above. Further analyses of the data showed that graduates with only post-Matric certificates or diplomas from historically black institutions that were suspected of providing inferior quality of education, and those graduating from the fields of business, commerce and management studies as well as education, training and development were more likely to be unemployed.

Dias and Posel (2006) looked at the relationship between education and unemployment by analysing the OHS 1995 and LFS 2003b data. Probit regressions were run to determine the probability of being unemployed. The results showed that individuals with tertiary education had a significantly lower probability of being unemployed, although these aggregate trends comprised different experiences among race groups and by gender. In addition, the increase in formally qualified labour was substantially larger than the increase in demand for skilled and semi-skilled labour over the period, and so unemployment rates increased even amongst graduates over the period. Finally, similar to what was found by Pauw et al. (2006), Dias and Posel argued that the prospective labour force participants did not specialise in specific fields of study (e.g., science and engineering as well as health and medical sciences) required by the labour market, or that employers were concerned about the quality of formal education.

To conclude, the studies above compared OHS 1995 with the latest available OHS or LFS at the time of writing to derive labour market ‘trends’. In general, it was found that the bulk of the increase of LF took place amongst blacks, females, those under 35 years, and those with some form of secondary education, while the majority of the increase of employment took place amongst blacks, females, people aged 25-54 years and those with at least Matric. Some studies argued that feminisation of labour market took place since the transition. In addition, the increase of real GDP, LF, LFPRs and employment were complemented by the increase of the number of unemployed and unemployment rates. Hence, jobless growth under the first definition (Altman 2003) did not take place, but jobless growth under the second definition (Altman 2003) did take place since the transition, mainly because employment, although showing an obvious increase since the transition, was insufficient to absorb the relatively rapid but expanding LF. Furthermore, unemployment rates amongst the black, female and relatively less educated people remained higher, but an alarming trend was the increase of the unemployment rate of people with post-Matric qualifications. Finally, the econometric analyses showed that blacks and males had a greater likelihood of joining the labour force, while whites, males, people aged 35 years and people with at least Matric were more likely to find employment.

Later in the chapter, all the labour surveys are taken into consideration to derive the real labour trends, and these trends are compared with the ‘trends’ derived from the studies above to determine if the findings are similar.

4.2.2 Review of studies that compared many surveys to derive labour market trends

As mentioned in Section 2.4.1, the sampling methodology and labour market status derivation methodology in OHS 1994-1995 were not known (due to the absence of the metadata document), so the results of the studies that compared OHS 1995 with another survey as discussed in Section 4.2.1 should be treated with caution. Few recent studies took this survey comparability issue into consideration and hence analysed many OHSs/LFSs available at the time of writing so as to derive detailed labour market trends for the period under investigation. Amongst these studies, Altman (2003 & 2008), and Casale, Muller and Posel (2004) paid particular attention to employment trends and the characteristics of the employed, while Arora and Ricci (2005) and Hlekiso & Mahlo (2009) focused on unemployment rates and unemployment by demographic characteristics respectively.

Altman (2003) used the 1995-1999 OHSs as well as LFS 2000b and LFS 2001b to explore whether South Africa was on a sustainable job-creating growth path, and she found that real GDP and the OHS/LFS non-agricultural formal sector employment moved in the same (upward) direction, but the relationship was not too strong. In addition, employment growth shifted from the formal sector to the informal sector, with the employment in the latter sector characterised by lower wages and fewer contractual benefits. In contrast, Altman’s other study (2008) used the 1995-1999 and all September LFSs in 2000-2006 to analyse employment trends, paying particular attention to whether job losses and gains recorded by OHSs reflected reality. She found that the plummeting and recovery of employment in the 1990s as reflected by the OHS employment data was questionable, due to reasons like coding errors, data inconsistencies (particularly in the agriculture, mining, as well as community, social and personal services industries) and the over-estimation of employment in OHS 1995.

Casale et al. (2004) used OHS 1995, OHS 1997, OHS 1999, LFS 2000b, LFS 2001b, LFS 2002b and LFS 2003a to analyse employment trends, focusing on whether the government’s claim that two million new jobs were created between 1995 and 2003 was valid. They found that the trend in employment was sensitive to the reference points used for the analysis and the increase of employment was likely to be inflated by changes in data capture and employment definitions over the years. In other words, the increase in employment was not real, bur rather an artefact of changing definitions and improved data collection. Furthermore, they argued that the real increase of employment might only have been approximately 1.4 million instead of two million. Finally, even if they were to accept that the increase of employment was indeed two million, more than half of the new jobs took place in subsistence farming, domestic work, and self-employment in the informal sector, with all types of work associated with very low returns. This finding was similar to Burger and Woolard's (2005).

As far as the studies that focused on unemployment trends are concerned, Arora and Ricci (2005) presented the narrow and broad black unemployment rates by age groups, educational attainment, area type, province and gender by considering all the OHSs and LFSs from 1995 to 2001. The results showed that unemployment rates were higher in younger age groups, amongst people with lower education, as well as in rural areas, Eastern Cape or KwaZulu-Natal provinces, and amongst females, confirming the findings by Bhorat and Oosthuizen (2005) and Oosthuizen (2006). On the other hand, the study by Hlekiso and Mahlo (2009), using all September LFSs in 2001-2007, first analysed the descriptive statistics on the narrow unemployed and found that most of them were females, blacks and aged 15-34 years, previously engaged in unskilled elementary occupations. The paper concluded by running a multinomial logistic regression on LFS 2007b to determine the likelihood of being inactive or unemployed, compared with the reference group (employed), and the results once again showed that blacks, females, people from the younger age cohorts and those with only Matric qualifications were more likely to be unemployed.

4.2.3 Review of studies that compared many surveys to focus on wage and discrimination trends

In addition to the studies discussed in Section 4.2.2, there were other studies that investigated most or all of the OHSs/LFSs available at the time of writing to derive the labour market trends in South Africa, but where the variables of interests were something other than LF, LFPR, employment and unemployment rates. Instead, the focus was on wage trends as well as whether employment and wage discrimination took place with the advent of democracy.

Brookes and Hinks (2004) used OHS 1995, OHS 1999, LFS 2000b, LFS 2001b and LFS 2002b to investigate the racial gap in employment likelihood of the broad labour force. Probit regressions were run to determine the likelihood of employment, before the racial employment gaps were divided into explained and unexplained components by means of

Oaxaca-Blinder decomposition[62]. The unexplained components were attributable to factors like differences in productivity of workers, differences in the quality of education, and discrimination. The results showed that the unexplained components of the white-black employment gap had increased from 12.0% to 17.4% between 1995 and 2002. The equivalent estimates for white-coloured and white-Indian employment gaps during the same period increased from 4.7% to 9.5% and from 0.9% to 5.5% respectively. Hence, the paper concluded that the employment hiring policies were ineffective in reducing the labour market inequalities of the past.

Burger and Jafta (2006), in addition to analysing the racial employment gap of the broad labour force as Brookes and Hinks did, also analysed the racial differential in occupational attainment and wages mainly by means of the Oaxaca and Blinder (1973) decomposition techniques[63]. The 1995-1999 OHSs and 2000-2004 LFSs data were used. Domestic workers, informal sector workers and people not in the working-age (15-65 years) were excluded from the analyses. In contrast to the findings by Brookes and Hinks (2004), Burger and Jafta found that the unexplained component of the white-black employment gap did not increase in the post-1994 period, but also did not show any obvious downward trend. The same observation was found in the white-black differential in skilled occupational attainment as well as wages. They concluded that affirmative action policies had no observable effect on the racial employment gap; their impact on the wage distribution was only limited to a small narrowing of wages at the top end of the wage distribution. In addition, there appeared to be a shift away from pure discrimination and towards differential returns to education. This was consistent with an increasingly important role for the quality of education in labour market outcomes.

The study by Shepherd (2008) used the 1995-1999 OHSs and all September LFSs from 2000 to 2006 to investigate gender discrimination in terms of wages amongst the blacks. The Oaxaca and Blinder decomposition technique was also adopted. Only formal sector workers (excluding subsistence agriculture, informal sector workers, self-employed and domestic workers) were included in the analysis. It was found that the driving factor behind an increasing and negative explained component of the gender wage gap amongst blacks was

[62] The Oaxaca-Blinder (1973) decomposition approach decomposes between-group (e.g., gender, race) average employment probability or wage gap into two parts: the explained component, which is due to the differences in factors like province, area type, educational attainment, years of experience, etc., and the unexplained component, which remains after controlling for these differences. In the case of wage gap, if productivity and educational quality differences could be completely controlled for, the unexplained component could be mainly explained by discrimination.

[63] Other decomposition techniques such as those of Brown, Moon and Zoloth (1980) as well as Juhn, Murphy and Pierce (1993) were also applied in their study.

improved distribution and returns to productive characteristics for women in certain occupations, and higher returns to education and employment in the public sector. However, black women were prevented from realizing this in the form of higher earnings, due to increasing levels of discrimination and returns to employment in certain industry categories for male workers.

Armstrong and Steenkamp (2008) used the 1995-1999 OHSs and all September LFSs from 2000 to 2005 to analyse the impact of union membership on wages, once again by using the Oaxaca and Blinder decomposition approach. They found that unions had an increasingly positive effect on members' wages and argued that this increase could be largely explained by the power trade unions had to increase the earnings of their members, not by differences in the attributes of unionized and non-unionized workers. Furthermore, unions were found to have "an inequality-reducing character, with union premiums for workers at the lower end of the wage distribution being greater than those for workers at the higher end of the distribution" (Armstrong and Steenkamp 2008: 2).

Finally, Burger and Yu (2006) used the 1995-1999 OHSs and all LFSs from 2000 to 2005 to derive wage trends in South Africa. They were concerned about the results of recent research (e.g., Leibbrandt, Levinsohn and McCrary 2005) that workers had, on average, experienced a considerable decrease in their real wage earnings in the post-apartheid period. However, they argued that this claim was based on choosing datasets on either side of Stats SA's changeover from OHS to LFS, which caused a discontinuously and inexplicably large decrease in mean earnings from the main job. After taking account of the inconsistencies in questionnaire design as well as the presence of outliers (e.g., the serious over-estimation of earnings of self-employed and informal sector workers in the OHSs, as well as respondents reporting zero or extremely high earnings from the main job), a fairly stable earnings series for formal sector employees was constructed. The data showed that real wage earnings increased slightly in the post-transition period. Looking at mean earnings trends by race, gender and skills category, they found that whites, males and those involved in highly-skilled occupations earned more on an average than non-whites, female and those involved in semi-skilled and unskilled occupations respectively, during the period under study.

In Sections 4.3-4.5, all the available OHSs, LFSs and QLFSs until the end of 2009 will be used to investigate detailed labour market trends since the transition, focusing on LF, LFPR, employment, work activities of the employed, unemployment and unemployment rates. In

Section 4.6, possible ways to further improve the comparability and reliability of labour market trends across the surveys as well as the critical evaluation of the changes (if any) of these trends after adopting these approaches will be looked at.

4.3 Characteristics of the labour force

This section studies the levels and trends of the LF by using OHS 1995-QLFS 2009Q4 data, as well as the demographic, location and educational attainment characteristics of the LF. The person weight variable is used to derive the weighted figures.

4.3.1 Number of labour force and labour force growth

Table 4.1 and Figure 4.1 present working-age population and LF size in 1994-2009. After a decline between OHS 1994 and OHS 1996, the LF in both narrow and broad terms showed a relatively large increase between OHS 1996 and LFS 2000a. The greatest increase occurred during the changeover from the OHS to the LFS, with an increase of more than 2 million in both narrow and broad terms. Since LFS 2000b, the LF in narrow terms stabilised at approximately 16 million, before increasing again from LFS 2005a onwards until reaching nearly 18 million in QLFS 2008Q4. A slight downward trend was observed since QLFS 2008Q2.

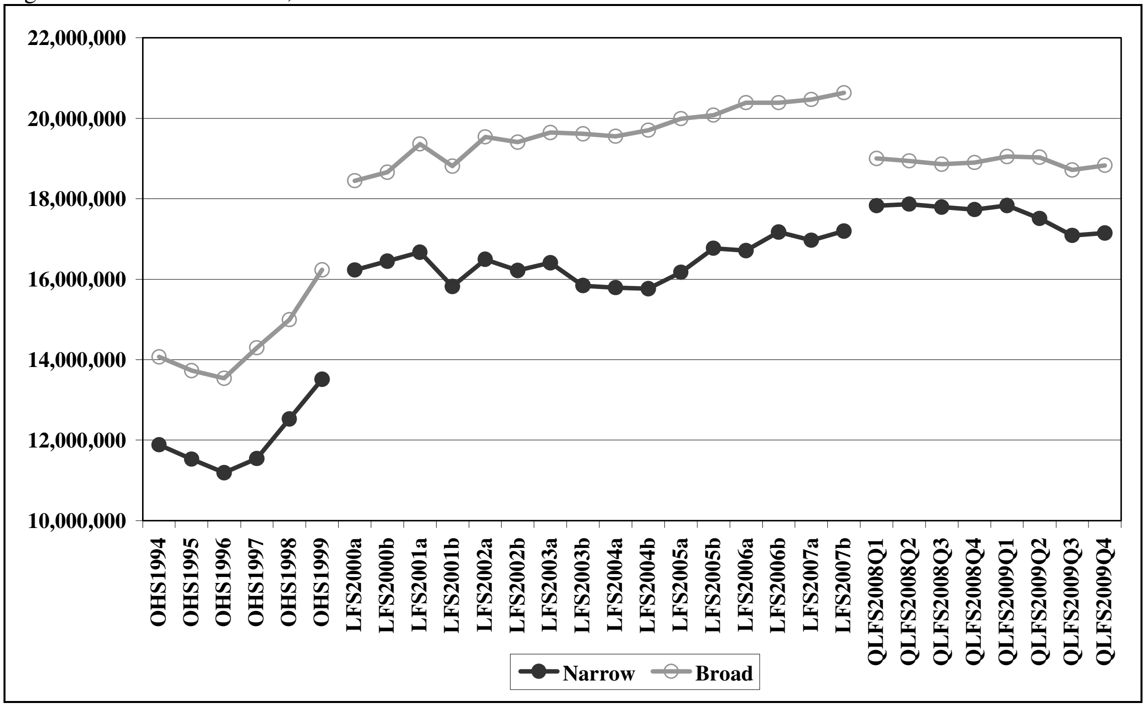
Table 4.1: The labour force (1 000s), 1994-2009

	Working-age population	Labour force – number		Labour force - % change	
		Narrow	Broad	Narrow	Broad
OHS 1994	24 075	11 884	14 073		
OHS 1995	24 191	11 528	13 731	-3.0%	-2.4%
OHS 1996	24 909	11 191	13 533	-2.9%	-1.4%
OHS 1997	25 506	11 544	14 296	3.2%	5.6%
OHS 1998	25 665	12 528	14 997	8.5%	4.9%
OHS 1999	26 247	13 510	16 231	7.8%	8.2%
LFS 2000a	26 465	16 206	18 424	20.1%	13.6%
LFS 2000b	27 836	16 381	18 596	1.3%	1.1%
LFS 2001a	28 062	16 668	19 361	1.4%	3.8%
LFS 2001b	28 084	15 817	18 808	-5.1%	-2.9%
LFS 2002a	28 298	16 494	19 535	4.3%	3.9%
LFS 2002b	28 495	16 215	19 405	-1.7%	-0.7%
LFS 2003a	28 725	16 409	19 642	1.2%	1.2%
LFS 2003b	28 906	15 841	19 610	-3.5%	-0.2%
LFS 2004a	29 100	15 788	19 550	-0.3%	-0.3%
LFS 2004b	29 271	15 761	19 704	-0.2%	0.8%
LFS 2005a	29 490	16 173	19 992	2.6%	1.5%
LFS 2005b	29 663	16 770	20 078	3.7%	0.4%
LFS 2006a	29 818	16 708	20 387	-0.4%	1.5%
LFS 2006b	29 973	17 173	20 386	2.8%	0.0%
LFS 2007a	30 161	16 966	20 465	-1.2%	0.4%
LFS 2007b	30 387	17 194	20 633	1.3%	0.8%

Table 4.1: Continued

	Working-age population	Labour force – number		Labour force - % change	
		Narrow	Broad	Narrow	Broad
QLFS 2008Q1	30 764	17 826	19 003	3.7%	-7.9%
QLFS 2008Q2	30 875	17 864	18 942	0.2%	-0.3%
QLFS 2008Q3	30 950	17 789	18 860	-0.4%	-0.4%
QLFS 2008Q4	31 047	17 733	18 901	-0.3%	0.2%
QLFS 2009Q1	31 145	17 833	19 048	0.6%	0.8%
QLFS 2009Q2	31 245	17 511	19 027	-1.8%	-0.1%
QLFS 2009Q3	31 325	17 086	18 717	-2.4%	-1.6%
QLFS 2009Q4	31 411	17 146	18 831	0.3%	0.6%

Figure 4.1: The labour force, 1994-2009

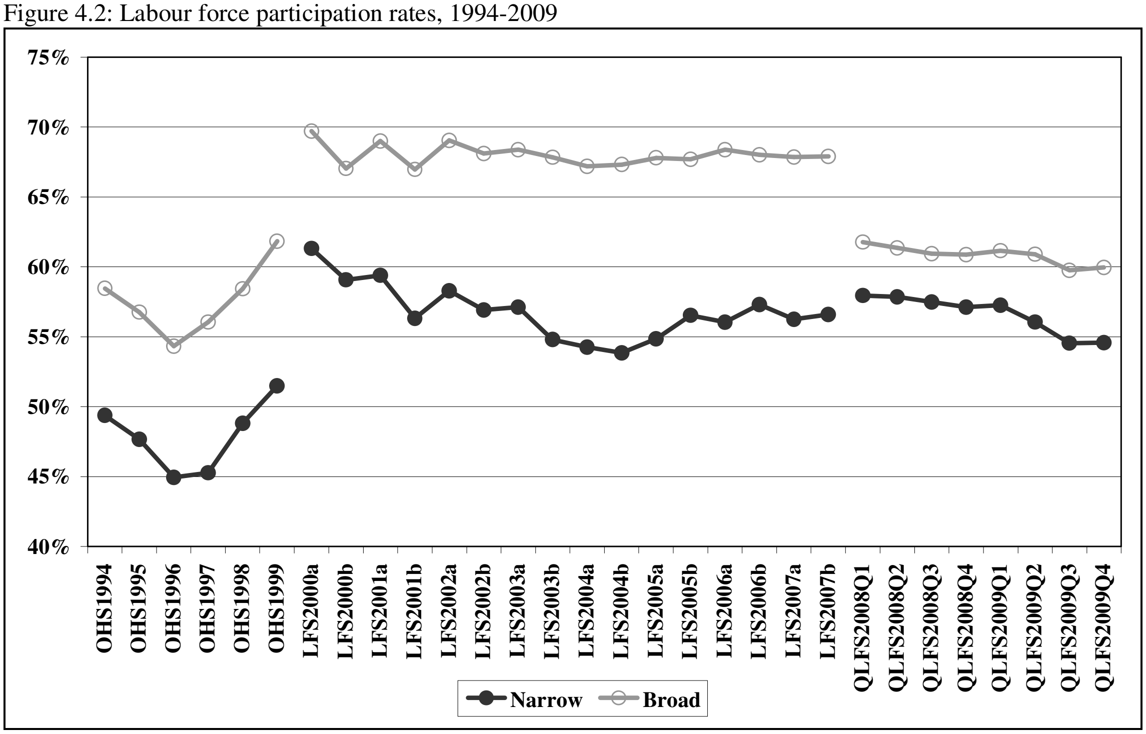


With regard to the LF in broad terms, between OHS 1994 and LFS 2000a, a similar trend was observed as in the narrow LF. In addition, the broad LF continued to increase between LFS 2000b and LFS 2007b, before an abrupt decline of 7.9% (1.63 million decrease) was observed between LFS 2007b and QLFS 2008Q1. This is probably due to the drastic differences in the broad labour market status derivation methodologies between LFS and QLFS, as mentioned in Section 2.4.2.2. During the QLFS period, the broad LF hovered around the 18.5-19.5 million ranges.

Figure 4.2 shows the trends in labour force participation rates (LFPRs). The narrow LFPR displays the same trend as the narrow LF, i.e., a decrease between OHS 1994 and OHS 1996, before a continuous upward trend was observed from OHS 1996 to LFS 2000a. Since then, the narrow LFPR showed a downward trend until LFS 2004b (decreasing from 61.3% to

53.8%), before a slight upward trend took place until QLFS 2008Q1. The narrow LFPR stayed within the 57%-58% range in all four 2008 QLFSs, before a downward trend was observed across the first three surveys in 2009. The broad LFPR showed the same trend as the narrow LFPR between OHS 1994 and LFS 2000a, and stayed within the 67%-69% range in all LFSs.

Figure 4.2: Labour force participation rates, 1994-2009



With regard to the possible reasons for the rapid increase of LF and LFPR in both narrow and broad terms between OHS 1997 and LFS 2000a (especially between OHS 1999 and LFS 2000a), it could be real[64], mainly due to the better capture of labour force as a result of the improvement of questionnaire structure, due to the different sampling techniques, or attributed to the changes in labour market status derivation methodology across the surveys. These issues, if possible, will be addressed later in Chapter 4 when the labour market trends are re-examined.

[64] For instance, the previously disadvantaged people like blacks and females felt more optimistic about their possible labour market outcomes as a result of the abolishment of the past discriminatory legislations since the transition, thereby deciding to enter the labour market to seek work. In addition, the study by Burger, Van der Berg and Von Fintel (2012) found that the rapid increase of labour force participation in the late 1990s could be due to the policies implemented by the South African Department of Education, which no longer allowed schools to accept students two years older than the correct grade-age. Hence, these students ended up entering the labour market for work (despite the fact that their employment prospects were low).

4.3.2 Demographic, geographic and educational attainment characteristics of the labour force

Table 4.2 shows the LF number by gender. The decrease of the LF in both narrow and broad terms between 1994 and 1996 was caused entirely by males (decreasing by about 200 000 in 1995 and 350 000 in 1996). In fact, the female LF was quite stable across these three OHSs. However, the abrupt increase of the LF between OHS 1999 and LFS 2000a mentioned earlier was more significant in both narrow and broad terms in the case of females. In addition, there were slight downward trends of the narrow LF number between LFS 2000b and LFS 2004b for both males and females. The broad LF of both genders increased steadily during the LFSs. Finally, the female share of the LF remained around 46% from LFS 2000b onwards in both narrow and broad terms. This result suggests that the period covered by LFS and QLFS showed no evidence of feminisation of the labour force, as claimed by Casale and Posel (2002), Casale (2004) and Van der Westhuizen et al. (2006).

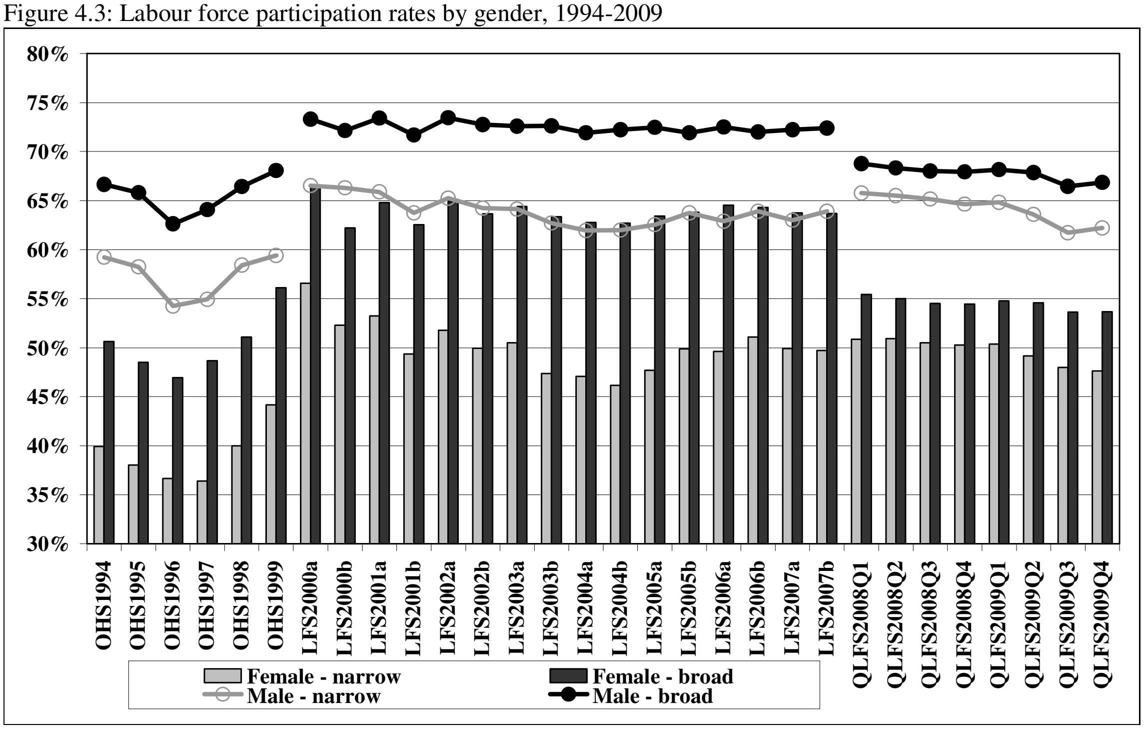
Table 4.2: Narrow and broad labour force by gender (1 000s), 1994-2009

	LF (1 000s)				Female share of LF	
	Male		Female			
	Narrow	Broad	Narrow	Broad	Narrow	Broad
OHS 1994	6 965	7 836	4 919	6 237	41.4%	44.3%
OHS 1995	6 713	7 587	4 815	6 144	41.8%	44.7%
OHS 1996	6 356	7 338	4 835	6 194	43.2%	45.8%
OHS 1997	6 708	7 825	4 837	6 471	41.9%	45.3%
OHS 1998	7 181	8 166	5 347	6 830	42.7%	45.5%
OHS 1999	7 479	8 571	6 023	7 651	44.6%	47.2%
LFS 2000a	8 385	9 239	7 816	9 180	48.2%	49.8%
LFS 2000b	8 916	9 703	7 465	8 892	45.6%	47.8%
LFS 2001a	8 988	10 016	7 677	9 342	46.1%	48.3%
LFS 2001b	8 668	9 750	7 150	9 058	45.2%	48.2%
LFS 2002a	8 926	10 050	7 567	9 485	45.9%	48.6%
LFS 2002b	8 921	10 105	7 289	9 295	45.0%	47.9%
LFS 2003a	8 953	10 132	7 454	9 508	45.4%	48.4%
LFS 2003b	8 770	10 155	7 071	9 455	44.6%	48.2%
LFS 2004a	8 710	10 114	7 073	9 431	44.8%	48.3%
LFS 2004b	8 791	10 239	6 961	9 455	44.2%	48.0%
LFS 2005a	8 899	10 311	7 267	9 671	45.0%	48.4%
LFS 2005b	9 103	10 270	7 661	9 799	45.7%	48.8%
LFS 2006a	9 057	10 440	7 649	9 945	45.8%	48.8%
LFS 2006b	9 277	10 449	7 896	9 937	46.0%	48.7%
LFS 2007a	9 205	10 551	7 760	9 913	45.7%	48.4%
LFS 2007b	9 378	10 624	7 805	9 997	45.4%	48.5%
QLFS 2008Q1	9 621	10 061	8 205	8 941	46.0%	47.1%
QLFS 2008Q2	9 622	10 038	8 242	8 904	46.1%	47.0%
QLFS 2008Q3	9 605	10 025	8 184	8 835	46.0%	46.8%
QLFS 2008Q4	9 561	10 048	8 172	8 853	46.1%	46.8%
QLFS 2009Q1	9 618	10 113	8 215	8 935	46.1%	46.9%
QLFS 2009Q2	9 467	10 104	8 044	8 924	45.9%	46.9%
QLFS 2009Q3	9 220	9 926	7 866	8 791	46.0%	47.0%
QLFS 2009Q4	9 324	10 020	7 822	8 811	45.6%	46.8%

* People with unspecified gender were excluded.

Figure 4.3 and Table A.1 in Appendix A present the LFPRs by gender. Both the narrow and broad rates for both genders decreased between 1994 and 1996, before showing a continuous upward trend between OHS 1997 and LFS 2000a. The abrupt increase between OHS 1999 and LFS 2000a took place in all four rates, but the increase was greatest in the female broad rate. There were slight downward trends of the narrow LFPR between LFS 2000b and LFS 2004b for both males and females, before they increased slightly and stabilised at about 65% and 50% respectively in the most recent surveys. The LFPR was very stable in the LFSs, hovering around 72% for males and 65% for females, before an abrupt decrease took place during the changeover from LFS to QLFS.

Figure 4.3: Labour force participation rates by gender, 1994-2009



The racial composition of the LF is presented in Table 4.3. The decrease of the LF in both narrow and broad terms between 1994 and 1995 was driven almost entirely by the white population group, while the decline between 1995 and 1996 was driven by the black population group. Additionally, the black share of the LF increased slightly throughout the period (even during the years covered by the LFS), while the white share became smaller. Figures 4.4 and 4.5 as well as Table A.2 in Appendix A show a similar pattern for the LFPRs of all four races (i.e., an increase during the years covered by the OHSs), although the increase in LFPR was more rapid for the black and Indian race groups. This was followed by a more abrupt increase during the changeover from the OHS to the LFS. The LFPRs became more stabilised in the LFSs.

Table 4.3: Narrow and broad labour force by race (1 000s), 1994-2009

	Black	Coloured	Indian	White	Black share#	White share#
	Narrow labour force (1 000s)					
OHS 1994	7 748	1 400	388	2 347	65.2%	19.8%
OHS 1995	7 829	1 362	401	1 936	67.9%	16.8%
OHS 1996	7 435	1 386	378	1 992	66.4%	17.8%
OHS 1997	7 842	1 370	401	1 932	67.9%	16.7%
OHS 1998	8 706	1 387	401	2 023	69.6%	16.2%
OHS 1999	9 408	1 517	464	2 101	69.7%	15.6%
LFS 2000a	11 869	1 656	493	2 204	73.2%	13.6%
LFS 2000b	12 027	1 637	488	2 261	73.3%	13.8%
LFS 2001a	12 268	1 677	492	2 208	73.7%	13.3%
LFS 2001b	11 421	1 620	528	2 230	72.3%	14.1%
LFS 2002a	11 999	1 728	508	2 239	72.8%	13.6%
LFS 2002b	11 806	1 677	540	2 173	72.9%	13.4%
LFS 2003a	11 962	1 723	530	2 184	72.9%	13.3%
LFS 2003b	11 454	1 659	519	2 200	72.3%	13.9%
LFS 2004a	11 455	1 694	503	2 127	72.6%	13.5%
LFS 2004b	11 454	1 657	484	2 130	72.8%	13.5%
LFS 2005a	11 821	1 691	515	2 119	73.2%	13.1%
LFS 2005b	12 403	1 712	523	2 097	74.1%	12.5%
LFS 2006a	12 356	1 711	484	2 138	74.0%	12.8%
LFS 2006b	12 769	1 750	499	2 100	74.6%	12.3%
LFS 2007a	12 671	1 747	473	2 048	74.8%	12.1%
LFS 2007b	12 763	1 722	510	2 163	74.4%	12.6%
QLFS 2008Q1	13 156	1 923	525	2 223	73.8%	12.5%
QLFS 2008Q2	13 248	1 910	534	2 172	74.2%	12.2%
QLFS 2008Q3	13 175	1 906	547	2 160	74.1%	12.1%
QLFS 2008Q4	13 149	1 914	531	2 139	74.1%	12.1%
QLFS 2009Q1	13 138	1 972	536	2 187	73.7%	12.3%
QLFS 2009Q2	12 882	1 940	521	2 167	73.6%	12.4%
QLFS 2009Q3	12 543	1 926	519	2 097	73.4%	12.3%
QLFS 2009Q4	12 566	1 948	506	2 125	73.3%	12.4%
	Black	**Coloured**	**Indian**	**White**	**Black share#**	**White share#**
	Broad labour force (1 000s)					
OHS 1994	9 791	1 476	403	2 403	69.6%	17.1%
OHS 1995	9 860	1 482	416	1 973	71.8%	14.4%
OHS 1996	9 621	1 494	396	2 022	71.1%	14.9%
OHS 1997	10 417	1 489	415	1 976	72.9%	13.8%
OHS 1998	10 959	1 534	425	2 067	73.1%	13.8%
OHS 1999	11 888	1 683	491	2 148	73.3%	13.2%
LFS 2000a	13 806	1 806	543	2 285	74.9%	12.4%
LFS 2000b	14 019	1 798	506	2 304	75.3%	12.4%
LFS 2001a	14 670	1 868	518	2 282	75.9%	11.8%
LFS 2001b	14 135	1 820	557	2 276	75.2%	12.1%
LFS 2002a	14 784	1 886	540	2 305	75.8%	11.8%
LFS 2002b	14 723	1 851	568	2 242	76.0%	11.6%
LFS 2003a	14 957	1 873	554	2 246	76.2%	11.4%
LFS 2003b	14 950	1 848	541	2 261	76.3%	11.5%
LFS 2004a	14 934	1 882	529	2 196	76.4%	11.2%
LFS 2004b	15 080	1 863	529	2 196	76.7%	11.2%
LFS 2005a	15 311	1 905	556	2 192	76.7%	11.0%
LFS 2005b	15 393	1 924	558	2 162	76.8%	10.8%
LFS 2006a	15 646	1 947	547	2 227	76.8%	10.9%
LFS 2006b	15 657	1 944	531	2 200	77.0%	10.8%
LFS 2007a	15 825	1 977	514	2 118	77.4%	10.4%
LFS 2007b	15 901	1 918	555	2 217	77.2%	10.8%

People whose race group was ‘other’ or ‘unspecified’ were excluded.

Figure 4.4: Narrow labour force participation rates by race, 1994-2009

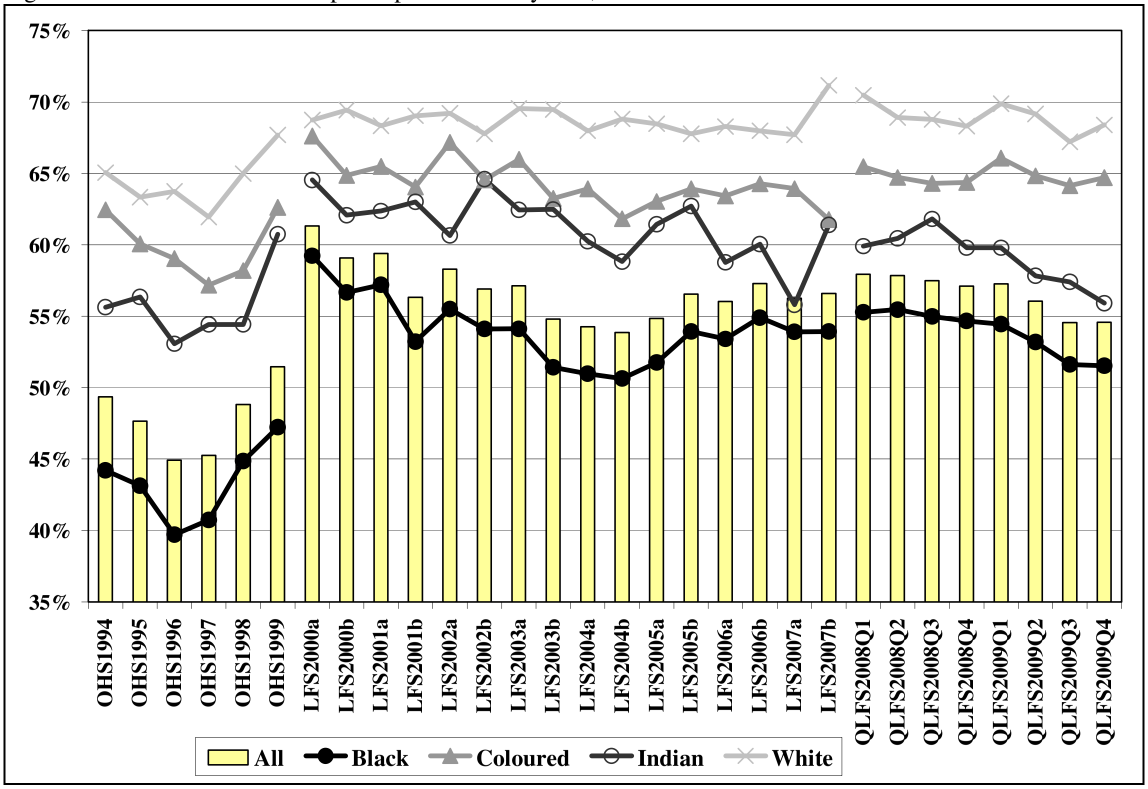
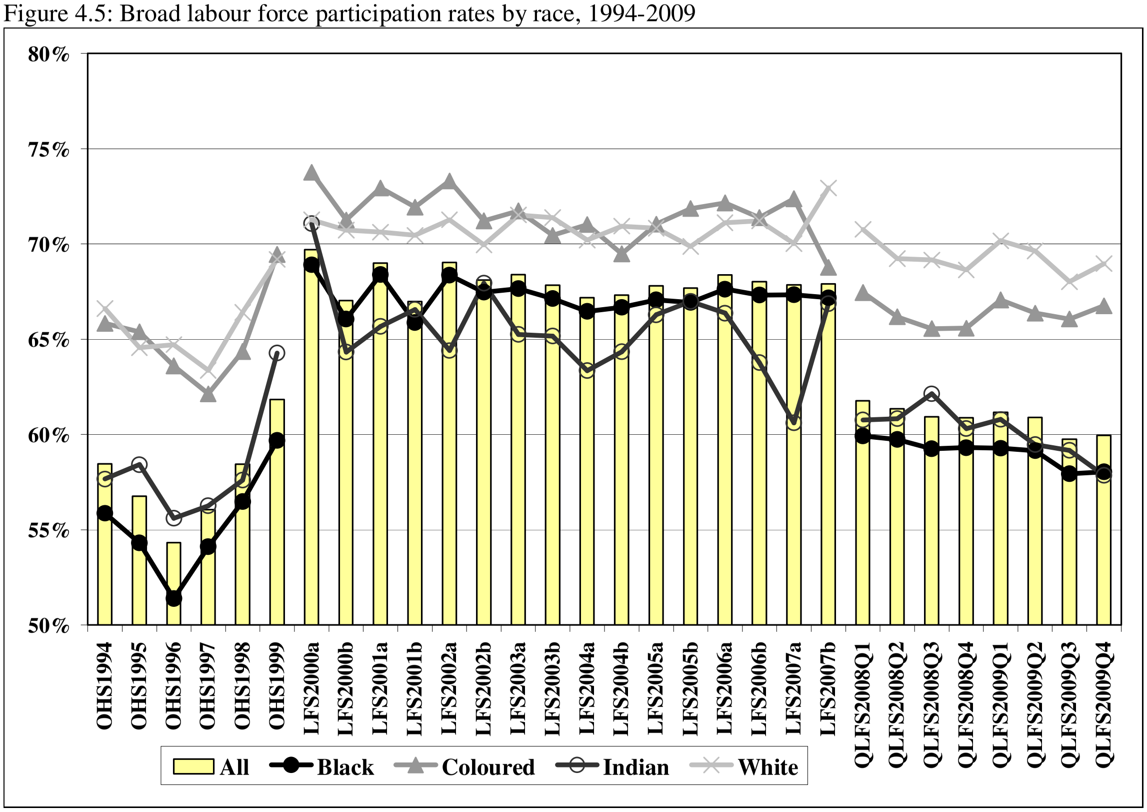
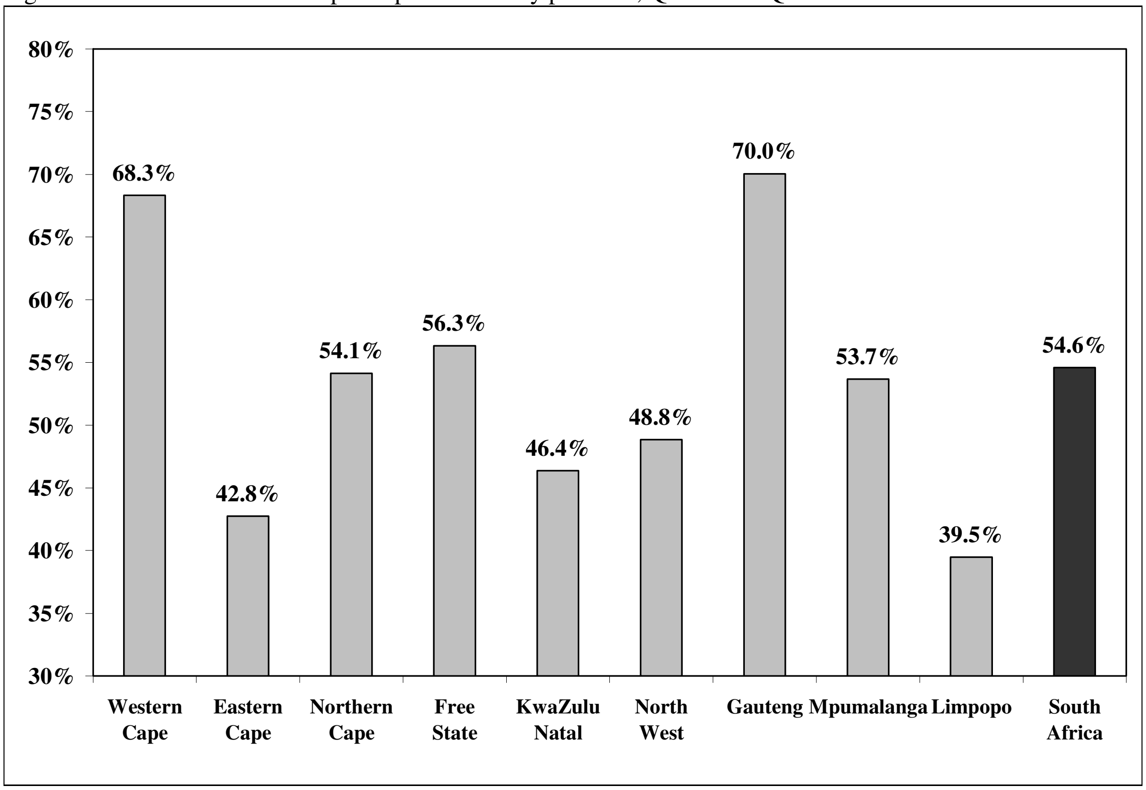


Figure 4.5: Broad labour force participation rates by race, 1994-2009



Looking at the LFPR by province, Tables A.3 and A.4 in Appendix A show that Western Cape, Free State and Gauteng were the only three provinces with LFPRs above the national rate during the period under study, in both narrow and broad terms. Gauteng was also the province showing the greatest increase of narrow LFPR between 1994 and 2009 (an increase of 10 percentage points), while in broad terms, Gauteng showed the fourth greatest increase in LFPR between 1994 and 2007 (an increase of nearly nine percentage points, while KwaZulu-Natal, Mpumalanga and Limpopo showed an increase of LFPR of more than 10 percentage points). Figure 4.6 shows the narrow LFPRs in QLFS 2009Q4. Regarding the provincial share of LF, Gauteng was the province with the greatest share (nearly 30% throughout the years in both terms).

Figure 4.6: Narrow labour force participation rates by province, QLFS 2009Q4



The LFPRs by age category is presented in Table 4.4. The LFPR was highest in the 25-34 year old and 35-44 year old age groups. The abrupt increase of LFPR between OHS 1999 and LFS 2000a was most rapid in the 55-65 year old age group (an increase of 14 percentage points in both narrow and broad terms), followed by the 15-24 year old age group (an increase of about 10 percentage points in both terms). As far as the share of LF by age category was concerned, the shares of each age category were very stable in both narrow and broad terms. The bulk of the LF (nearly 60%) was between the age of 25 and 44 years.

Table 4.4: Labour force participation rates by age category, 1994-2009

	Narrow LFPRs					Broad LFPRs				
	15-24 years	**25-34 years**	**35-44 years**	**45-54 years**	**55-65 years**	**15-24 years**	**25-34 years**	**35-44 years**	**45-54 years**	**55-65 years**
OHS 1994	23.5%	66.2%	69.9%	62.9%	33.6%	30.4%	79.2%	80.6%	71.3%	37.3%
OHS 1995	21.7%	63.9%	69.6%	62.7%	31.7%	29.4%	77.5%	79.1%	69.9%	34.6%
OHS 1996	20.3%	59.9%	66.9%	57.8%	30.6%	27.6%	74.6%	77.3%	65.4%	33.1%
OHS 1997	19.7%	60.4%	66.2%	58.2%	30.5%	27.8%	77.0%	78.1%	66.9%	34.0%
OHS 1998	22.9%	66.4%	70.2%	62.6%	31.1%	31.1%	80.6%	80.1%	69.8%	34.6%
OHS 1999	25.4%	68.8%	73.4%	65.1%	34.0%	34.7%	83.7%	84.2%	72.6%	37.3%
LFS 2000a	35.5%	78.4%	81.2%	75.0%	48.0%	44.0%	89.9%	88.6%	80.9%	51.4%
LFS 2000b	32.1%	76.2%	79.5%	72.8%	47.0%	40.8%	87.0%	86.6%	78.2%	49.7%
LFS 2001a	32.5%	76.9%	80.1%	72.0%	46.6%	43.6%	89.4%	88.3%	78.3%	50.0%
LFS 2001b	30.5%	75.1%	76.8%	67.2%	39.2%	42.6%	88.6%	86.4%	75.2%	42.6%
LFS 2002a	33.1%	75.7%	78.5%	70.0%	41.7%	45.1%	90.2%	87.7%	77.0%	45.6%
LFS 2002b	31.3%	75.4%	77.4%	67.6%	39.3%	43.6%	89.5%	87.6%	76.5%	43.4%
LFS 2003a	31.4%	76.0%	77.7%	68.1%	38.3%	44.5%	90.2%	87.1%	76.3%	42.8%
LFS 2003b	29.3%	73.1%	74.5%	66.4%	37.2%	44.5%	90.2%	86.0%	74.7%	40.8%
LFS 2004a	28.8%	72.5%	73.7%	65.3%	37.6%	43.7%	89.2%	85.3%	73.9%	42.0%
LFS 2004b	28.2%	71.3%	73.3%	66.3%	37.9%	42.9%	89.6%	86.1%	75.1%	41.6%
LFS 2005a	28.2%	73.1%	74.2%	66.7%	41.0%	42.9%	90.1%	86.1%	75.4%	45.4%
LFS 2005b	30.7%	74.8%	75.4%	69.1%	40.3%	42.8%	89.9%	86.3%	76.5%	43.2%
LFS 2006a	29.9%	74.5%	74.4%	68.2%	41.3%	43.5%	90.4%	86.3%	76.9%	45.4%
LFS 2006b	30.6%	75.1%	77.4%	70.1%	41.8%	41.9%	89.8%	87.8%	77.2%	45.1%
LFS 2007a	30.1%	75.0%	75.4%	67.2%	40.5%	42.5%	89.7%	86.9%	76.3%	44.8%
LFS 2007b	29.7%	74.6%	76.9%	69.3%	40.4%	42.2%	89.2%	87.7%	77.6%	44.4%
QLFS 2008Q1	31.1%	76.4%	78.4%	69.6%	42.0%	35.1%	81.4%	81.9%	72.6%	43.5%
QLFS 2008Q2	30.8%	76.6%	77.9%	69.5%	42.3%	34.4%	81.5%	81.0%	72.2%	43.6%
QLFS 2008Q3	30.2%	76.2%	78.2%	69.2%	41.6%	33.7%	81.1%	81.2%	71.9%	42.9%
QLFS 2008Q4	29.4%	76.2%	77.6%	69.6%	41.1%	33.2%	81.3%	81.3%	72.3%	42.6%
QLFS 2009Q1	29.7%	75.8%	77.7%	70.2%	41.7%	33.9%	81.1%	81.2%	72.9%	43.2%
QLFS 2009Q2	28.6%	74.2%	77.2%	69.0%	40.3%	33.8%	80.8%	81.8%	72.2%	41.6%
QLFS 2009Q3	26.4%	72.9%	76.6%	67.3%	38.8%	31.9%	80.0%	81.8%	70.9%	40.3%
QLFS 2009Q4	26.8%	73.0%	76.0%	67.5%	38.7%	32.1%	80.4%	81.2%	71.9%	40.3%

The educational attainment[65] of the LF declined in both the number and the share of people with no or incomplete primary schooling, which coincided with the increase in both the number and share of people with at least Matric. The results (in narrow terms) are presented in Table 4.5. The labour force has gradually become more educated on average[66]. In addition, Figure 4.7 as well as Tables A.5 and A.6 in Appendix A show the LFPRs in each educational attainment category, and the LFPRs are higher in the more educated categories. Furthermore, the abrupt increase of the LFPR between OHS 1999 and LFS 2000a was more substantial in the groups in which people had lower levels of educational attainment.

[65] The question to capture the highest educational attainment of the respondents was asked very consistently in all surveys under study, as there was one question that asked the survey participant to declare the highest level of education completed, and the educational attainment categories were very similar across the surveys. The only exceptions were in OHS 1997 and OHS 1998. In these two surveys, the respondents were asked two questions, namely the highest school standard passed and the highest completed tertiary education (Yu 2007: 15 & 44).

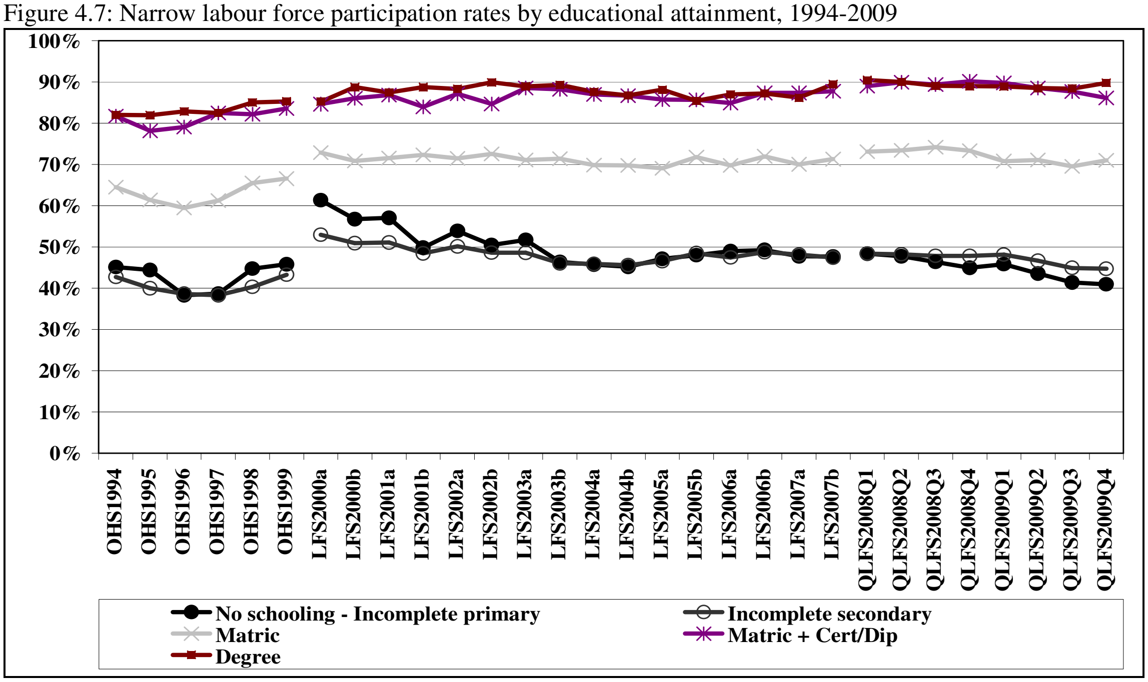
[66] Figure A.5 in Appendix B provides more detail by showing the proportion of the LF with at least Matric by race, and it can be seen that this proportion is the highest for whites but lowest for blacks. However, the latter share nearly doubled between OHS 1994 and QLFS 2009Q4, increasing to about 40% in the latter survey.

Table 4.5: Narrow labour force by educational attainment (1 000s), 1994-2009

	No schooling	Incomplete primary	Incomplete secondary	Matric	Matric + Cert/Dip	Degree	% with at least Matric#
OHS 1994	942	2 016	5 144	2 439	832	468	31.6%
OHS 1995	912	1 913	4 682	2 533	939	456	34.4%
OHS 1996	879	1 752	4 524	2 587	809	521	35.4%
OHS 1997	925	1 705	4 822	2 679	914	464	35.2%
OHS 1998	1 060	2 065	4 990	2 968	967	448	35.1%
OHS 1999	927	2 269	5 293	3 188	847	709	35.9%
LFS 2000a	1 214	2 896	6 726	3 539	999	655	32.4%
LFS 2000b	1 203	2 930	6 702	3 412	1 133	891	33.4%
LFS 2001a	1 179	2 797	6 823	3 813	1 112	795	34.6%
LFS 2001b	1 001	2 569	6 412	3 840	1 043	792	36.2%
LFS 2002a	1 105	2 575	6 753	4 009	1 100	810	36.2%
LFS 2002b	988	2 464	6 634	4 033	1 104	837	37.2%
LFS 2003a	948	2 498	6 710	4 149	1 136	847	37.6%
LFS 2003b	816	2 256	6 356	4 327	1 172	824	40.1%
LFS 2004a	820	2 178	6 355	4 432	1 104	836	40.5%
LFS 2004b	846	2 095	6 439	4 341	1 112	777	39.9%
LFS 2005a	786	2 131	6 665	4 498	1 164	836	40.4%
LFS 2005b	866	2 135	6 972	4 661	1 208	813	40.1%
LFS 2006a	818	2 100	6 845	4 757	1 304	822	41.3%
LFS 2006b	822	2 069	7 195	4 838	1 370	790	41.0%
LFS 2007a	765	1 975	7 199	4 805	1 349	799	41.2%
LFS 2007b	790	2 027	7 042	4 660	1 485	1 082	42.3%
QLFS 2008Q1	679	1 910	7 396	5 148	1 601	901	43.4%
QLFS 2008Q2	702	1 869	7 380	5 190	1 584	925	43.6%
QLFS 2008Q3	676	1 841	7 358	5 184	1 600	944	43.9%
QLFS 2008Q4	646	1 768	7 415	5 098	1 676	938	44.0%
QLFS 2009Q1	629	1 737	7 385	5 206	1 750	963	44.8%
QLFS 2009Q2	589	1 642	7 225	5 214	1 738	947	45.5%
QLFS 2009Q3	548	1 533	6 969	5 144	1 753	942	46.4%
QLFS 2009Q4	532	1 529	6 993	5 293	1 675	945	46.6%

People with unspecified educational attainment were excluded.

Figure 4.7: Narrow labour force participation rates by educational attainment, 1994-2009



In summary, the LFPR decreased between OHS 1994 and OHS 1996, before increasing until OHS 1999. An abrupt increase was observed between OHS 1999 and LFS 2000, after which it appeared to stabilise. In addition, using all available labour surveys, it was found that the increase of labour force participation was relatively more rapid in the case of blacks and those in the younger age cohorts. Furthermore, the labour force became more educated throughout the years. These findings are similar to those derived by comparing the OHS 1995 with the latest available LFS (see Section 4.2.1). However, although the female labour force participation increased since the transition, there was no indication that feminisation of the labour force took place after examining all surveys. This result contradicts what was found by studies that only compared OHS 1995 with an LFS (Section 4.2.1).

4.3.2 Multivariate analyses on labour force participation likelihood

Since the analyses in Section 4.3.1 only took one or two variables into account, multivariate analyses are conducted here, by means of probit regressions[67] on labour force participation likelihood in both narrow and broad terms. The explanatory variables are as follows:

- Province dummy variables (Reference group: Eastern Cape)
- Race dummy variables (Reference group: Black)
- Gender (Reference group: Female)
- Age category dummy variables (Reference group: 15-24 years)
- Educational attainment spline variables: No education to Grade 6, Grade 7 to Grade 11
- Educational attainment dummy variables: Matric, Matric plus Certificate or Diploma, Degree or above
- Household head dummy variable
- A dummy variable indicating the participant was married or living with a partner
- Number of children aged 0-14 years in the household
- Number of elderly aged 60 years or above in the household
- Number of males aged 15-59 years in the household
- Number of females aged 15-59 years in the household
- Survey dummy variables (Reference group: OHS 1994)

The results are presented in Table 4.6 and Figure A.6 in Appendix A. The following people are more likely to be labour force participants in both narrow and broad terms: Coloured

[67] In an Ordinary Least Squares regression, the dependent variable is a continuous variable (e.g., earnings, age in years, price of a good, etc.), while in a probit regression, the dependent variable is a binary variable with a value of either zero or one (e.g., labour force participation status, with zero and one meaning ‘no’ and ‘yes’ respectively).

males, aged 25 years or above, married household heads, those with higher educational attainment, households with fewer children or elderly but greater number of males and females aged 15-59 years, and those staying in provinces other than Eastern Cape and Limpopo. In particular, the positive male dummy coefficient once again implies that feminisation of labour force did not take place since the transition.

The survey dummy variables indicate that the labour force participation likelihood is lower in OHS 1996-1998 but higher in other surveys, compared with OHS 1994. The LFS 2000a dummy has the greatest coefficient in absolute terms (0.3615 in the narrow probit regression and 0.3625 in the broad narrow probit regression), and this could be attributable to abrupt increase of LFPR between OHS 1999 and LFS 2000a. Finally, under the broad definition, labour force participation probability is lower in the QLFSs, as indicated by the negative sign of the coefficients of the QLFS dummies. This is due to the abrupt decrease of labour force number in the QLFSs due to the adoption of the stricter labour market status derivation methodology under the broad definition in these surveys (refer to Section 3.11).

Table 4.6: Probit regressions on labour force participation

	Narrow definition		**Broad definition**	
	Coefficient	**Absolute value of Z-statistic**	**Coefficient**	**Absolute value of Z-statistic**
Western Cape	0.3011	1349.78	0.1866	814.18
Northern Cape	0.1289	357.18	0.1014	273.96
Free State	0.2035	885.38	0.1357	575.35
KwaZulu-Natal	0.1112	652.67	0.0709	411.41
North West	0.0450	209.64	0.0828	377.8
Gauteng	0.3312	1926.85	0.2999	1689.64
Mpumalanga	0.1709	759.72	0.1271	552.47
Limpopo	-0.1890	944.53	-0.0872	436.45
Coloured	0.1574	751.86	0.0495	228.23
Indian	-0.1527	502.29	-0.3685	1187.1
White	-0.1866	984.89	-0.3958	2019.11
Male	0.3538	3089.47	0.3094	2628.34
No education to Grade 6	0.0164	456.50	0.0215	590.72
Grade 7 to Grade 11	0.0360	953.75	0.0216	564.79
Matric	0.4668	3071.54	0.5537	3448.17
Matric + Certificate/Diploma	0.8829	3090.07	0.8516	2760.64
Degree	0.8427	2294.39	0.8021	2083.58
Age: 25-34 years	0.9652	7131.10	1.1741	8182.11
Age: 35-44 years	0.9875	5851.97	1.0763	6072.95
Age: 45-54 years	0.7096	3711.66	0.6843	3494.51
Age: 55-65 years	0.0193	86.20	-0.1246	552.74
Household head	0.5596	4127.61	0.5183	3624.64
Married or living together with a partner	0.1947	1629.67	0.1914	1514.97

Table 4.6: Continued

	Narrow definition		Broad definition	
	Coefficient	Absolute value of Z-statistic	Coefficient	Absolute value of Z-statistic
Number of children 0-14 years	-0.0523	1594.35	-0.0347	1052.59
Number of elderly 60 years or above	-0.1158	1252.77	-0.0815	883.34
Number of male 15-59 years	0.0006	13.50	0.0043	91.54
Number of female 15-59 years	0.0303	655.92	0.0398	845.05
OHS 1995	-0.0596	147.26	-0.0653	160.04
OHS 1996	-0.1463	364.13	-0.1411	349.36
OHS 1997	-0.1414	354.46	-0.0969	241.3
OHS 1998	-0.0296	74.40	-0.0150	37.36
OHS 1999	0.0338	84.96	0.0827	204.86
LFS 2000a	0.3615	910.62	0.3625	888.19
LFS 2000b	0.2721	692.77	0.2532	631.78
LFS 2001a	0.2749	701.52	0.3202	796.17
LFS 2001b	0.1823	466.29	0.2539	634.98
LFS 2002a	0.2302	589.92	0.3146	784.23
LFS 2002b	0.1844	472.96	0.2828	706.74
LFS 2003a	0.1819	467.33	0.2822	706.35
LFS 2003b	0.0967	248.79	0.2606	653.15
LFS 2004a	0.0762	196.67	0.2281	574.68
LFS 2004b	0.0532	137.38	0.2315	582.28
LFS 2005a	0.0816	211.11	0.2418	609.13
LFS 2005b	0.1367	353.70	0.2370	597.71
LFS 2006a	0.1174	304.64	0.2587	653.08
LFS 2006b	0.1482	383.72	0.2403	606.3
LFS 2007a	0.1107	287.59	0.2296	580.97
LFS 2007b	0.1142	296.23	0.2257	571.14
QLFS 2008Q1	0.1518	394.59	-0.0083	21.44
QLFS 2008Q2	0.1490	387.28	-0.0221	56.84
QLFS 2008Q3	0.1354	352.14	-0.0359	92.36
QLFS 2008Q4	0.1195	310.87	-0.0406	104.62
QLFS 2009Q1	0.1135	295.62	-0.0436	112.29
QLFS 2009Q2	0.0790	205.85	-0.0453	116.77
QLFS 2009Q3	0.0229	59.74	-0.0889	229.39
QLFS 2009Q4	0.0297	77.46	-0.0783	202.11
Constant	-1.3275	3428.73	-1.0474	2680.39
Observed probability	0.5510		0.6408	
Number of observations (weighted)	854 986 437		854 986 437	
Chi	2.829e+08		2.747e+08	
Pseudo R-squared	0.2405		0.2460	

Note: All explanatory variables are statistically significant at 1%.

4.4 Employment

This section studies the levels and trends of employment by using OHS 1995-QLFS 2009Q4 data, as well as the demographic, location and educational attainment characteristics of the employment. Whether jobless growth took place since the transition or not will also be looked at. The person weight variable is used to derive the weighted figures.

4.4.1 Number of employed and employment growth

Figure 4.8 and Table 4.7 show the number of employed and its trend between consecutive surveys. The employment figures fluctuated substantially throughout the period under investigation. The greater number of employed in OHS 1994-1995 compared with other OHS years was mainly the result of greater employment in the agriculture, forestry, fishing and hunting industry in these two surveys. An abrupt increase of nearly one million took place in OHS 1999, followed by an even greater increase of about 1.52 million in LFS 2000a.

The abovementioned rapid increase of employment between OHS 1998 and LFS 2000a could either be real, due to the better capturing of employed with the improvement of questionnaire (see Section 3.11), or due to the different weighting techniques across the surveys (see Section 2.4.1). Section 4.6 will address these issues, if possible, to re-examine the employment trends.

Figure 4.8: Number of employed, 1994-2009

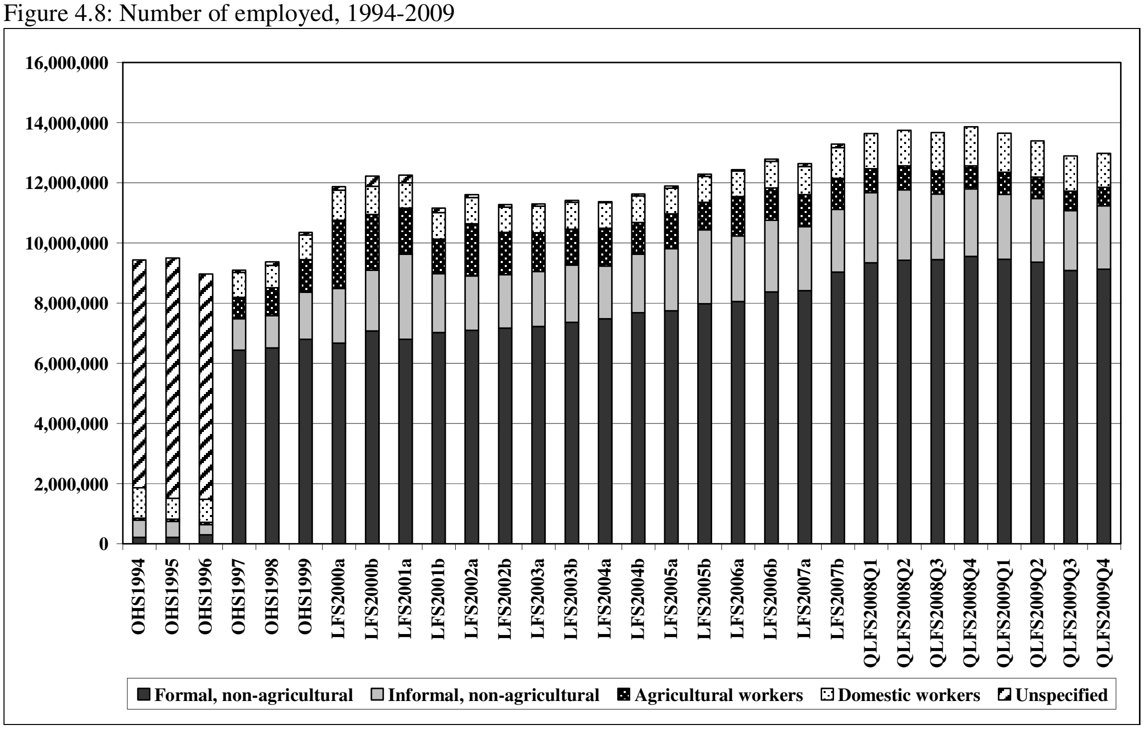


Table 4.7: Number of employed (1 000s) and the employment change between consecutive surveys, 1994-2009

	Number of employed	Change in the number of employed	
		Absolute	Percentage
OHS 1994	9 439		
OHS 1995	9 499	60	0.6%
OHS 1996	8 966	-533	-5.6%
OHS 1997	9 094	127	1.4%
OHS 1998	9 370	276	3.0%
OHS 1999	10 356	986	10.5%
LFS 2000a	11 874	1 518	14.7%
LFS 2000b	12 224	350	2.9%
LFS 2001a	12 260	36	0.3%
LFS 2001b	11 168	-1 093	-8.9%
LFS 2002a	11 603	436	3.9%
LFS 2002b	11 284	-319	-2.8%
LFS 2003a	11 298	14	0.1%
LFS 2003b	11 411	114	1.0%
LFS 2004a	11 378	-33	-0.3%
LFS 2004b	11 630	252	2.2%
LFS 2005a	11 894	264	2.3%
LFS 2005b	12 288	394	3.3%
LFS 2006a	12 438	150	1.2%
LFS 2006b	12 787	349	2.8%
LFS 2007a	12 635	-152	-1.2%
LFS 2007b	13 293	658	5.2%
QLFS 2008Q1	13 637	344	2.6%
QLFS 2008Q2	13 749	112	0.8%
QLFS 2008Q3	13 669	-81	-0.6%
QLFS 2008Q4	13 862	193	1.4%
QLFS 2009Q1	13 653	-209	-1.5%
QLFS 2009Q2	13 388	-264	-1.9%
QLFS 2009Q3	12 897	-491	-3.7%
QLFS 2009Q4	12 984	87	0.7%

A substantial decrease of employment by 1.09 million was recorded in LFS 2001b, before employment showed an upward trend in general until QLFS 2008Q4, reaching nearly 14 million. However, employment declined continuously by approximately one million between QLFS 2008Q4 and QLFS 2009Q3, mainly as a result of the global economic recession.

As Tables 4.1 and 4.7 above have presented the labour force and employment numbers between 1994 and 2009, the question arises as whether the economy has generated employment at a pace that could absorb the net labour force entrants. To answer this question, target growth ratio (TGR), actual growth ratio (AGR) and employment absorption ratio (EAR) need to be derived:

- The target growth ratio (TRG) measures how fast employment would have had to expand in order to provide work for all the net entrants to the labour market from period

X to period Y. Period X and Y need not be two consecutive years. $TGR = \frac{LF_Y - LF_X}{E_X}$, where LF and E stand for the number of labour force and employed respectively.

- The actual growth ratio (AGR) is the growth rate of the number of employed from period X to period Y. $AGR = \frac{E_Y - E_X}{E_X}$.
- The employment absorption ratio (EAR) measures the proportion of the net increase in the labour force from period X to period Y that finds employment during the same period. $EAR = \frac{E_Y - E_X}{LF_Y - LF_X} = \frac{AGR}{TGR}$. An EAR of 100% means that the full net increase in the labour force between the two periods was employed.

Serious care needs to be taken when deciding which two surveys to choose for calculating these three ratios, as the selection of surveys for comparison may lead to very different results. Table 4.8 presents two examples: when comparing OHS 1995 with LFS 2006b, the EARs were quite low, except for Indians and whites. However, a comparison between LFS 2006b and LFS 2001b indicates that the economy seemed to have created more jobs than required in narrow terms (EAR equaled 119.4%), even for blacks (EAR equaled 113.5%). One could therefore argue that the economy created more than enough employment opportunities between LFS 2001b and LFS 2006b.

Table 4.8: Employment performance of the economy, LFS 2006b vs. OHS 1995 and LFS 2006b vs. LFS 2001b

	Narrow			**Broad**		
	LFS 2006b vs. OHS 1995					
	TGR	**AGR**	**EAR**	**TGR**	**AGR**	**EAR**
Black	80.5%	44.6%	55.4%	94.5%	44.6%	47.2%
Coloured	33.9%	23.2%	68.3%	40.3%	23.2%	57.4%
Indian	27.5%	25.9%	94.3%	32.0%	25.9%	80.9%
White	8.9%	7.8%	88.5%	12.2%	7.8%	64.3%
Male	44.3%	26.3%	59.4%	49.4%	26.3%	53.2%
Female	83.0%	47.6%	57.3%	102.2%	47.6%	46.5%
All	59.4%	34.6%	58.2%	70.1%	34.6%	49.4%
	LFS 2006b vs. LFS 2001b					
	TGR	**AGR**	**EAR**	**TGR**	**AGR**	**EAR**
Black	18.3%	20.8%	113.5%	20.7%	20.8%	100.4%
Coloured	10.2%	10.4%	102.5%	9.7%	10.4%	107.0%
Indian	-6.6%	5.4%	-82.2%	-6.2%	5.4%	-86.6%
White	-6.2%	-4.5%	72.5%	-3.6%	-4.5%	123.9%
Male	9.5%	13.6%	144.0%	10.9%	13.6%	125.6%
Female	15.8%	15.7%	99.4%	18.6%	15.7%	84.4%
All	12.1%	14.5%	119.4%	14.1%	14.5%	102.6%

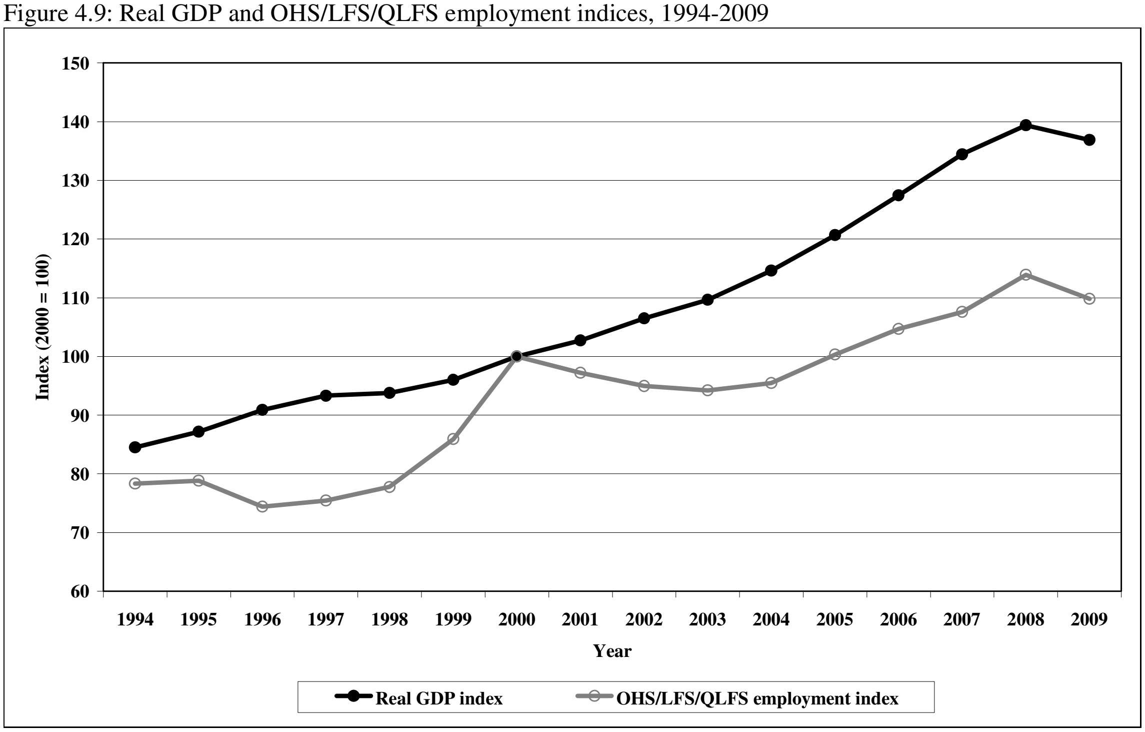
Table 4.9 provides more information by showing the TGRs, AGRs and EARs in narrow terms when comparing two surveys selected from different years, and the results show that the EAR was the highest (78.8%) when comparing LFS 2004b with QLFS 2009Q4.

Table 4.9: Target growth ratios, actual growth ratios and employment absorption ratios during various periods, in narrow terms

	TGR	AGR	EAR
OHS 1994 vs. OHS 1999	17.2%	9.7%	56.4%
OHS 1999 vs. LFS 2004b	21.7%	12.3%	56.6%
LFS 2004b vs. QLFS 2009Q4	14.9%	11.8%	78.8%
OHS 1995 vs. LFS 2005b	55.2%	29.4%	53.2%
OHS 1995 vs. LFS 2007b	59.7%	39.9%	67.0%
OHS 1995 vs. QLFS 2009Q4	59.1%	36.7%	62.0%

As the results of the analyses in Tables 4.8 and 4.9 suggest that the employment growth was not rapid enough to absorb all the net labour force entrants since the advent of democracy until 2009, the next question to examine is whether jobless growth took place or not since the transition. Figure 4.9 shows that, despite the fluctuations in some years, employment displayed an increasing trend in general (except between 2002 and 2003) while real GDP increased continuously. Hence, jobless growth under the first definition did not take place.

Figure 4.9: Real GDP and OHS/LFS/QLFS employment indices, 1994-2009



Sources: South African Reserve Bank Website and own calculations using OHS/LFS/QLFS data.
Note: Real GDP (2000 prices) – Code: KBP6006Y (Real GDP is converted into an index, and the index in 2000 is equal to 100)
Note: OHS/LFS/QLFS employment index was derived by taking the average of the figures from the surveys taking place in the same year, i.e., the LFS 2000a and LFS 2000b employment numbers were averaged to derive the 2000 index value, and so forth.

When unemployment is analysed in Section 4.5, it will be seen that unemployment rates in both narrow and broad terms increased continuously until LFS 2003a, and the second interpretation of jobless growth (real GDP increase complemented by rising unemployment rate) took place in the South African economy until early 2003.

With regard to whether employment took place in the formal or informal sectors, Table 4.10 shows whether the employed were involved in formal or informal employment[68]. The large number of unspecified people in OHS 1994-1996 (See Figure 4.8) was due to the fact that employees were not asked to declare their formal/informal sector status[69]. Moreover, the aforementioned abrupt increase of the number of employed in OHS 1999 and LFS 2000a could be explained by the sudden increase of the number of the informal workers (increasing from 1.08 million in OHS 1998 to 1.57 million in OHS 1999) and subsistence agriculture workers (increasing from 0.28 million in OHS 1999 to 1.51 million in LFS 2000a). This could be attributed to the better capturing of informal sector employment as a result of the improvement of questionnaire design (See Table 3.6). In other words, it is possible that informal sector employment was under-estimated in OHS 1997-1999.

The informal sector employment was the highest in LFS 2001a (2.84 million). This is due to the fact that more probing questions were asked about self-employment and small businesses in a follow-up survey. Hence, this might have led to a larger number of respondents than usual classifying themselves as informal sector workers[70]. The sudden decrease of informal sector employment between LFS 2001a and LFS 2001b (from 2.84 million to 1.97 million) could explain the rapid decline of the number of employed between the two surveys (See Table 2.7). In contrast, the big increase of formal sector employment between LFS 2007a and LFS 2007b (from 8.42 million to 9.03 million) primarily resulted in the increase of the number of employed between these two surveys (See Table 2.7). Finally, since LFS 2001b, informal sector employment as percentage of total non-agricultural employment (excluding domestic workers) has stabilised at approximately 20%.

---

[68] Until LFS 2007b, Stats SA focused on whether an enterprise that the employed worked for (regardless of whether they worked as employees or self-employed) was registered or not, by using the respondents' answer to the direct, self-perception question on the enterprise registration status (Essop & Yu, 2008a: 6-7). With the inception of QLFS in 2008, the enterprise characteristics remained the important criteria to distinguish informal workers, but questions other than the direct question mentioned above were also considered. The self-employed were considered as informal workers if their businesses were not registered for either income tax or value-added tax, while the employees were classified as informal sector workers if they were not registered for income tax and worked in establishments that employed fewer than five employees (Essop & Yu, 2008b: 4-5).

[69] Only the self-employed were asked to declare the enterprise registration status in these three OHSs.

[70] Devey, Skinner and Valodia (2006: 309) argue that the LFS 2001a informal employment figure is not an outlier, but rather the 'correct' estimate, with the estimates in other surveys significantly under-representing the true level of informal employment.

Table 4.10: Formal vs. Informal employment (1 000s), 1994-2009

	[A]	[B]	[C]	[D]	[E]	[F]	[G]	All
OHS 1994	1 023	568	218	15	45	0	7 570	9 439
OHS 1995	695	522	219	27	50	0	7 987	9 499
OHS 1996	766	330	304	25	56	0	7 485	8 966
OHS 1997	828	1 043	6 436	187	526	0	73	9 094
OHS 1998	747	1 077	6 508	202	725	0	110	9 370
OHS 1999	812	1 572	6 796	284	799	0	93	10 356
LFS 2000a	1 003	1 820	6 673	1 508	757	86	29	11 874
LFS 2000b	941	2 026	7 077	1 074	767	108	230	12 224
LFS 2001a	844	2 836	6 798	742	785	214	40	12 260
LFS 2001b	881	1 965	7 019	382	765	127	29	11 168
LFS 2002a	875	1 821	7 089	863	865	75	15	11 603
LFS 2002b	843	1 779	7 173	550	852	62	26	11 284
LFS 2003a	885	1 828	7 223	443	841	57	19	11 298
LFS 2003b	895	1 901	7 365	365	832	36	18	11 411
LFS 2004a	846	1 765	7 474	341	913	26	15	11 378
LFS 2004b	880	1 944	7 685	425	624	53	19	11 630
LFS 2005a	849	2 068	7 742	513	647	28	47	11 894
LFS 2005b	858	2 460	7 980	338	578	34	41	12 288
LFS 2006a	849	2 188	8 052	703	606	14	27	12 438
LFS 2006b	885	2 376	8 376	473	605	47	25	12 787
LFS 2007a	936	2 129	8 415	460	603	53	40	12 635
LFS 2007b	1 024	2 084	9 034	368	667	47	69	13 293
QLFS 2008Q1	1 165	2 325	9 344	161	642	0	0	13 637
QLFS 2008Q2	1 186	2 348	9 424	122	670	0	0	13 749
QLFS 2008Q3	1 274	2 179	9 449	111	656	0	0	13 669
QLFS 2008Q4	1 299	2 250	9 550	121	643	0	0	13 862
QLFS 2009Q1	1 300	2 157	9 457	128	610	0	0	13 653
QLFS 2009Q2	1 195	2 114	9 368	98	614	0	0	13 388
QLFS 2009Q3	1 166	1 996	9 081	76	578	0	0	12 897
QLFS 2009Q4	1 135	2 110	9 123	105	510	0	0	12 984

[A]: Domestic workers
[B]: Informal sector workers (excluding agricultural workers)
[C]: Formal sector workers (excluding agricultural workers)
[D]: Agricultural workers in informal sector (also known as subsistence agricultural workers)
[E]: Agricultural workers in formal sector (also known as commercial agricultural workers)
[F]: Don't know
[G]: Not specified

Table 4.11 presents the number and the proportion of employed working as employees and self-employed. It is apparent that self-employment was under-estimated during the OHS years, as a result of the problematic categorization of the question that resulted in the under-capture of low-income, informally self-employed workers (See Table 3.6). Apart from the over-estimation of subsistence agriculture workers mentioned earlier, the doubling of the number of self-employed between OHS 1999 and LFS 2000a could also lead to the rapid increase in the number of employed in the latter survey. The unusually large decline in the number of employed in LFS 20001b seemed to be mainly caused by a decrease of the number of self-employed (most of them were informal sector workers), which declined from 3.22 million to 2.14 million. Finally, employees as percentage of all employed stabilised at the

80%-85% ranges from LFS 2001b onwards.

Table 4.11: Type of employment, 1994-2009

	Employee		Self-Employed	
	Number (1 000s)	Percentage#	Number (1 000s)	Percentage#
OHS 1994	8 018	84.9%	1 421	15.1%
OHS 1995	8 123	85.5%	1 376	14.5%
OHS 1996	8 313	93.2%	611	6.8%
OHS 1997	8 167	89.8%	926	10.2%
OHS 1998	8 340	89.0%	1 026	11.0%
OHS 1999	8 845	85.5%	1 506	14.5%
LFS 2000a	8 787	74.1%	3 074	25.9%
LFS 2000b	9 371	76.8%	2 825	23.2%
LFS 2001a	9 025	73.7%	3 218	26.3%
LFS 2001b	9 012	80.8%	2 144	19.2%
LFS 2002a	9 082	78.4%	2 509	21.6%
LFS 2002b	9 082	80.6%	2 191	19.4%
LFS 2003a	9 194	81.4%	2 099	18.6%
LFS 2003b	9 276	81.3%	2 131	18.7%
LFS 2004a	9 356	82.3%	2 019	17.7%
LFS 2004b	9 414	81.0%	2 207	19.0%
LFS 2005a	9 536	80.3%	2 340	19.7%
LFS 2005b	9 846	80.3%	2 423	19.7%
LFS 2006a	9 772	78.6%	2 659	21.4%
LFS 2006b	10 184	79.7%	2 593	20.3%
LFS 2007a	10 253	81.3%	2 365	18.7%
LFS 2007b	10 936	82.5%	2 323	17.5%
QLFS 2008Q1	11 520	84.5%	2 117	15.5%
QLFS 2008Q2	11 589	84.3%	2 161	15.7%
QLFS 2008Q3	11 532	84.4%	2 136	15.6%
QLFS 2008Q4	11 677	84.2%	2 185	15.8%
QLFS 2009Q1	11 498	84.2%	2 155	15.8%
QLFS 2009Q2	11 301	84.4%	2 087	15.6%
QLFS 2009Q3	11 007	85.4%	1 889	14.6%
QLFS 2009Q4	10 978	84.6%	2 006	15.4%

People whose employment type could not be specified were excluded.

4.4.2 Demographic, geographic and educational attainment characteristics of the employed

Table 4.12 presents employment trends by gender. The figures for females were relatively more erratic, even during the LFS years. The abrupt increase in the number of employed between OHS 1999 and LFS 2000a as mentioned before was greater for females (an increase of more than 1 million or 28.2%), which caused the female share of the employed to increase by 5 percentage points to 47% over the same period. Subsequently, the female share stabilised at about 42%. In the QLFSs, the female share of employed slightly increased to about 45%. Hence, the increase of female share of employed between 39.0% in OHS 1994 to 44.6% in QLFS 2009Q4 suggests that feminisation of employment took place to a certain extent since the transition (as suggested by the studies that only compared two surveys, as mentioned in Section 4.2.1), but the male share is still more dominant (greater than 50%) in all surveys.

Table 4.12: Employment trends by gender, 1994-2009

	Number of employed (1 000s)		Percentage change		Share of employed#	
	Male	Female	Male	Female	Male	Female
OHS 1994	5 759	3 680			61.0%	39.0%
OHS 1995	5 789	3 710	0.5%	0.8%	60.9%	39.1%
OHS 1996	5 327	3 639	-8.0%	-1.9%	59.4%	40.6%
OHS 1997	5 539	3 555	4.0%	-2.3%	60.9%	39.1%
OHS 1998	5 635	3 736	1.7%	5.1%	60.1%	39.9%
OHS 1999	6 001	4 348	6.5%	16.4%	58.0%	42.0%
LFS 2000a	6 295	5 574	4.9%	28.2%	53.0%	47.0%
LFS 2000b	6 935	5 289	10.2%	-5.1%	56.7%	43.3%
LFS 2001a	6 780	5 478	-2.2%	3.6%	55.3%	44.7%
LFS 2001b	6 435	4 733	-5.1%	-13.6%	57.6%	42.4%
LFS 2002a	6 598	5 005	2.5%	5.7%	56.9%	43.1%
LFS 2002b	6 607	4 673	0.1%	-6.6%	58.6%	41.4%
LFS 2003a	6 517	4 779	-1.4%	2.3%	57.7%	42.3%
LFS 2003b	6 607	4 805	1.4%	0.5%	57.9%	42.1%
LFS 2004a	6 632	4 747	0.4%	-1.2%	58.3%	41.7%
LFS 2004b	6 765	4 860	2.0%	2.4%	58.2%	41.8%
LFS 2005a	6 904	4 985	2.1%	2.6%	58.1%	41.9%
LFS 2005b	7 048	5 236	2.1%	5.0%	57.4%	42.6%
LFS 2006a	7 104	5 333	0.8%	1.9%	57.1%	42.9%
LFS 2006b	7 313	5 474	2.9%	2.6%	57.2%	42.8%
LFS 2007a	7 263	5 372	-0.7%	-1.9%	57.5%	42.5%
LFS 2007b	7 518	5 768	3.5%	7.4%	56.6%	43.4%
QLFS 2008Q1	7 639	5 998	1.6%	4.0%	56.0%	44.0%
QLFS 2008Q2	7 710	6 039	0.9%	0.7%	56.1%	43.9%
QLFS 2008Q3	7 633	6 035	-1.0%	-0.1%	55.8%	44.2%
QLFS 2008Q4	7 758	6 104	1.6%	1.1%	56.0%	44.0%
QLFS 2009Q1	7 584	6 069	-2.2%	-0.6%	55.5%	44.5%
QLFS 2009Q2	7 407	5 981	-2.3%	-1.4%	55.3%	44.7%
QLFS 2009Q3	7 109	5 788	-4.0%	-3.2%	55.1%	44.9%
QLFS 2009Q4	7 194	5 790	1.2%	0.1%	55.4%	44.6%

People with unspecified gender were excluded.

Table 4.13 shows that the bulk of the net increase in employment took place amongst blacks. The slight increase in the black share of the employed was complemented by the small decrease in the white share. (However, the black share has stabilised at 69%-70% since LFS 2005b.) The abrupt increase of black employment between OHS 1994 and OHS 1995 took place at the expense of white employment. In addition, the abrupt increase of employment between OHS 1999 and LFS 2000a as well as between LFS 2007a and LFS 2007b, and the sudden decline of employment between LFS 2001a and LFS 2001b (See Table 2.7) were almost entirely the result of the rapid changes of employment amongst blacks during the same period. Furthermore, black employment increased by more than two million between LFS 2003b and QLFS 2008Q4 (i.e., before retrenchment took place in 2009), while white employment remained at around two million during the same period.

Table 4.13: Employment trends by race, 1994-2009

	Number of employed (1 000s)				Share of employed#			
	Black	**Coloured**	**Indian**	**White**	**Black**	**Coloured**	**Indian**	**White**
OHS 1994	5 730	1 129	333	2 248	60.7%	12.0%	3.5%	23.8%
OHS 1995	6 136	1 145	359	1 860	64.6%	12.1%	3.8%	19.6%
OHS 1996	5 489	1 222	337	1 918	61.2%	13.6%	3.8%	21.4%
OHS 1997	5 714	1 161	362	1 857	62.8%	12.8%	4.0%	20.4%
OHS 1998	5 915	1 168	342	1 934	63.2%	12.5%	3.7%	20.7%
OHS 1999	6 660	1 286	392	2 002	64.4%	12.4%	3.8%	19.4%
LFS 2000a	8 120	1 317	395	2 036	68.4%	11.1%	3.3%	17.2%
LFS 2000b	8 363	1 333	408	2 096	68.6%	10.9%	3.3%	17.2%
LFS 2001a	8 456	1 321	410	2 056	69.1%	10.8%	3.3%	16.8%
LFS 2001b	7 344	1 277	428	2 100	65.9%	11.5%	3.8%	18.8%
LFS 2002a	7 777	1 312	406	2 093	67.1%	11.3%	3.5%	18.1%
LFS 2002b	7 507	1 292	429	2 043	66.6%	11.5%	3.8%	18.1%
LFS 2003a	7 498	1 338	411	2 042	66.4%	11.8%	3.6%	18.1%
LFS 2003b	7 571	1 309	433	2 090	66.4%	11.5%	3.8%	18.3%
LFS 2004a	7 540	1 388	420	2 023	66.3%	12.2%	3.7%	17.8%
LFS 2004b	7 866	1 296	419	2 015	67.8%	11.2%	3.6%	17.4%
LFS 2005a	8 080	1 356	423	2 012	68.1%	11.4%	3.6%	16.9%
LFS 2005b	8 498	1 328	440	1 991	69.3%	10.8%	3.6%	16.2%
LFS 2006a	8 568	1 387	430	2 037	69.0%	11.2%	3.5%	16.4%
LFS 2006b	8 874	1 410	451	2 006	69.6%	11.1%	3.5%	15.7%
LFS 2007a	8 841	1 401	408	1 960	70.1%	11.1%	3.2%	15.5%
LFS 2007b	9 343	1 367	468	2 080	70.5%	10.3%	3.5%	15.7%
QLFS 2008Q1	9 513	1 555	463	2 106	69.8%	11.4%	3.4%	15.4%
QLFS 2008Q2	9 672	1 538	466	2 073	70.3%	11.2%	3.4%	15.1%
QLFS 2008Q3	9 573	1 541	483	2 071	70.0%	11.3%	3.5%	15.1%
QLFS 2008Q4	9 746	1 571	469	2 076	70.3%	11.3%	3.4%	15.0%
QLFS 2009Q1	9 509	1 588	468	2 088	69.6%	11.6%	3.4%	15.3%
QLFS 2009Q2	9 297	1 562	462	2 068	69.4%	11.7%	3.5%	15.4%
QLFS 2009Q3	8 938	1 510	454	1 996	69.3%	11.7%	3.5%	15.5%
QLFS 2009Q4	8 969	1 543	450	2 022	69.1%	11.9%	3.5%	15.6%

People whose race group was ‘other’ or ‘unspecified’ were excluded.

Employment has been consistently concentrated in Western Cape, Gauteng and KwaZulu-Natal. These three provinces accounted for about 60% of the total throughout the period under consideration. The provincial shares have been very stable throughout the years (See Table A.7 in Appendix A), with the exception of a slight increase in the share of Gauteng and a slight decrease in the share of the Free State. Figure 4.10 shows the provincial shares of employment in QLFS 2009Q4.

Figure 4.10: Provincial shares of employed, QLFS 2009Q4

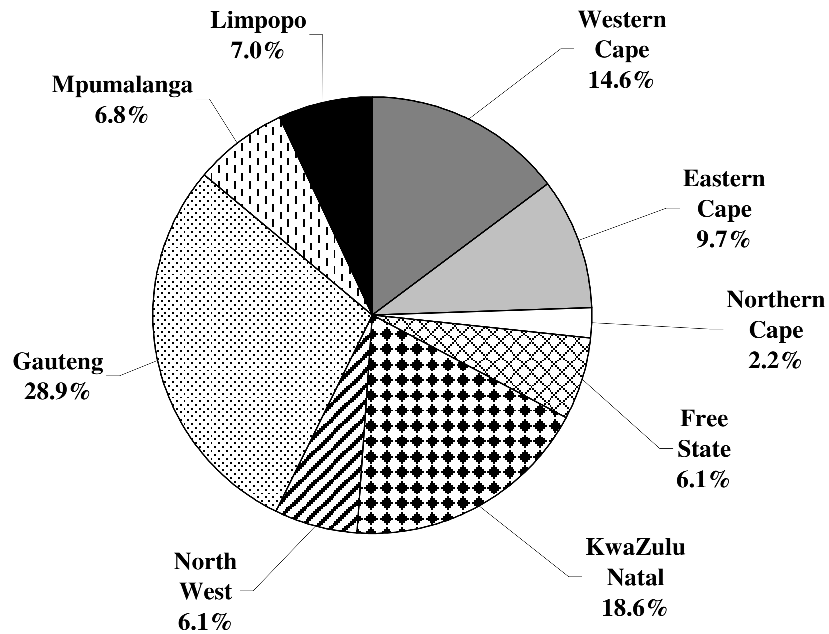
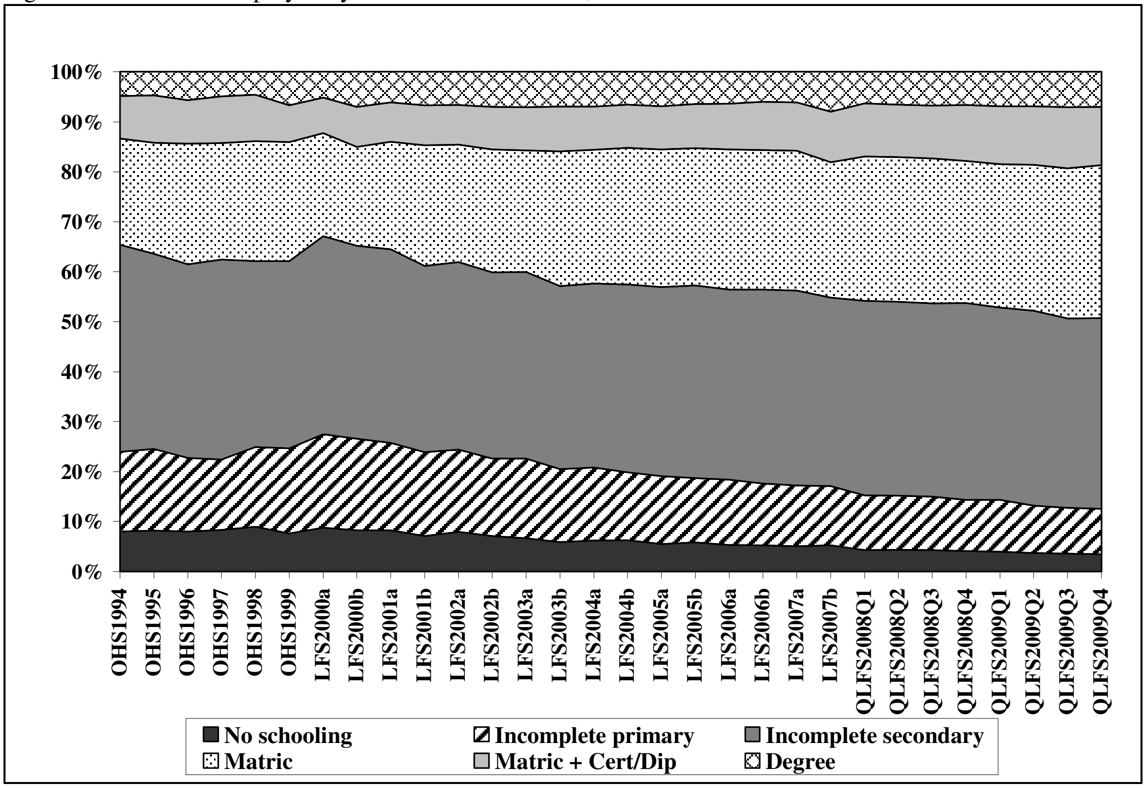


Table A.8 in Appendix A shows that the 25-34 year old and 35-44 year old age groups consistently accounted for about 60% of total employment during the period under study. Slightly above 10% of the employed were aged 15-24 years, while approximately 9% of the employed were aged 55-65 years.

Finally, as far as the employment by educational attainment was concerned, Figure 4.11 and Table A.9 in Appendix A indicate that the employed have become more educated on average, as the share of employed with at least Matric displayed a continuous upward trend, increasing from approximately 35% in OHS 1994 to nearly 50% in QLFS 2009Q4. Figure A.7 of Appendix A provides more information by showing the percentage of employed with at least Matric by race, and it can be seen that this proportion increased in all races during the period under study. The increase was the greatest in the black employed, as this proportion doubled from 20% in OHS 1994 to 40% in QLFS 2009Q4.

Figure 4.11: Share of employed by educational attainment, 1994-2009



Note: The employed with unspecified educational attainment were excluded.

In summary, the employment trends discussed indicate that black employment showed a more rapid increase and that there is no strong indication of feminisation of employment. In addition, most of the employed resided in Western Cape, KwaZulu-Natal or Gauteng and were aged 25-44 years. The bulk of employment creation benefited the black population, and the employed became more educated on average, as indicated by the increase of the share of employed with at least Matric. This share almost doubled between 1994 and 2009 among black employed.

To conclude Section 4.4.2, after examining all labour surveys, it was found that job creation increased more rapidly amongst the black females, aged 25-44 years at the time of the survey, with at least Matric.

4.4.3 Multivariate analyses on employment likelihood

Since the analyses on employment in Section 4.4.2 only took one or two variables into account, multivariate analyses are conducted here to investigate the factors influencing the labour force's likelihood of employment. However, as mentioned before (See footnote 61), the labour force is not a random sample, but has rather undergone a selection process whereby some of the working-age population decided to enter the labour force while others opted not.

Consequently, the estimated results of a simple probit regression on employment would suffer from sample selection bias.

In order to address this problem, a Heckman two-step approach is adopted. First, the labour force participation probit equation, as done in Table 4.6, is estimated. This equation allows the estimation of the inverse Mills ratio (i.e., lambda), which is in turn included in the employment probit, making the latter regression conditional on labour force participation. The explanatory variables included for the regression are province, race, gender, age category, educational attainment, household head status and marital status.

The results of the employment probit regressions (for both narrow and broad LF) are presented in Table 4.14 and Figure A.8 in Appendix A). First, the coefficients for lambda in both regressions are statistically significant. This indicates that the labour force indeed differ from their counterparts, who decided not to participate in the labour force. Thus, the two-step Heckman approach is necessary.

Table 4.14: Heckprobit regressions on employment

	Narrow labour force		Broad labour force	
	Coefficient	**Absolute value of Z-statistic**	**Coefficient**	**Absolute value of Z-statistic**
Western Cape	0.1608	465.48	0.3517	1186.96
Northern Cape	-0.0524	98.94	0.0519	110.66
Free State	-0.0210	60.76	0.1335	448.12
KwaZulu-Natal	0.0623	236.33	0.1307	580.39
North West	-0.0014	4.45	0.0272	98.12
Gauteng	-0.0278	98.39	0.1520	640.40
Mpumalanga	0.0539	157.80	0.1515	516.97
Limpopo	-0.1117	342.99	-0.2173	-819.56
Coloured	0.3666	1254.07	0.4263	1627.73
Indian	0.4966	1074.18	0.5265	1205.56
White	0.9264	2868.19	0.8968	2934.91
Male	0.0562	293.65	0.1994	1290.41
No education to Grade 6	-0.0191	323.24	-0.0030	60.88
Grade 7 to Grade 11	-0.0182	320.82	0.0078	160.95
Matric	0.1661	625.63	0.2777	1131.34
Matric + Certificate/Diploma	0.6100	1363.42	0.8459	2174.38
Degree	0.8500	1401.49	1.0728	1960.91
Age: 25-34 years	0.2822	622.63	0.5282	1158.00
Age: 35-44 years	0.4664	980.26	0.7152	1589.95
Age: 45-54 years	0.5804	1332.39	0.7726	2003.04
Age: 55-65 years	0.8004	1980.24	0.7818	2300.73
Household head	0.6639	2476.29	0.7809	3751.12
Married or living together with a partner	0.3272	1941.07	0.3294	2261.09

Table 4.14: Continued

	Narrow definition		Broad definition	
	Coefficient	Absolute value of Z-statistic	Coefficient	Absolute value of Z-statistic
OHS 1995	0.1403	211.43	0.0811	146.92
OHS 1996	0.0304	46.10	-0.0443	80.54
OHS 1997	-0.0098	15.11	-0.0998	185.37
OHS 1998	-0.1495	240.96	-0.1210	228.34
OHS 1999	-0.0903	146.24	-0.0692	132.04
LFS 2000a	-0.1349	225.01	0.0503	96.90
LFS 2000b	-0.1308	219.46	0.0400	77.65
LFS 2001a	-0.1619	273.52	-0.0170	33.24
LFS 2001b	-0.2718	460.95	-0.1501	293.96
LFS 2002a	-0.2753	469.65	-0.1378	270.74
LFS 2002b	-0.3031	517.91	-0.1869	367.99
LFS 2003a	-0.3315	568.46	-0.2061	406.61
LFS 2003b	-0.2448	415.66	-0.1908	376.30
LFS 2004a	-0.2320	393.85	-0.1893	374.47
LFS 2004b	-0.1848	312.43	-0.1751	346.84
LFS 2005a	-0.1888	320.96	-0.1606	318.60
LFS 2005b	-0.1768	302.32	-0.1008	200.15
LFS 2006a	-0.1247	212.48	-0.0953	189.47
LFS 2006b	-0.1285	219.78	-0.0633	125.81
LFS 2007a	-0.1281	218.96	-0.0927	184.87
LFS 2007b	-0.0349	59.19	-0.0144	28.72
QLFS 2008Q1	-0.0571	97.74	0.1417	276.43
QLFS 2008Q2	-0.0317	54.16	0.1738	337.60
QLFS 2008Q3	-0.0415	70.89	0.1624	315.17
QLFS 2008Q4	0.0069	11.70	0.1900	367.68
QLFS 2009Q1	-0.0612	104.76	0.1287	250.51
QLFS 2009Q2	-0.0733	125.50	0.0817	159.62
QLFS 2009Q3	-0.1241	212.54	0.0097	18.99
QLFS 2009Q4	-0.1016	173.89	0.0236	46.20
Lambda	-0.0439	59.32	0.3211	442.37
Constant	-0.0448	33.76	-1.1856	1112.15
Observed probability	0.7477		0.6429	
Number of observations (weighted)	471 091 048		471 091 048	
Chi	9.887e+07		1.524e+08	
Pseudo R-squared	0.1858		0.2134	

Note: All explanatory variables are statistically significant at 1%.

The following people are more likely to be employed in both regressions: non-black males, those aged 25 years or above, married household heads, those with at least Matric, and those staying in Western Cape, KwaZulu-Natal and Mpumalanga provinces (under the broad definition, the labour force from any other province other than Eastern Cape and Limpopo is more likely to be employed). The survey dummy variables indicate that employment likelihood is lower in OHS 1996-1998 but higher in others, compared with OHS 1994.

Furthermore, in the Heckprobit regression for the narrow LF, employment likelihood is greater in OHS 1995, OHS 1996 and QLFS 2008Q4, as indicated by the positive coefficient of these dummy variables. Finally, in the Heckprobit probit for the broad LF, employment probability is greater in OHS 1995, LFS 2000a and LFS 2000b, as well as all QLFSs.

4.4.4 Work activities of the employed

Although employment increased in the South African economy between 1996 and 2008 in general, the experiences in various occupations and industries differed. Table 4.15 presents the percentage of employed in each broad occupation category. The skilled agricultural and fishery worker category (column F) showed the biggest fluctuations. In fact, the rapid increase in the number of employed in LFS 2000a and the decrease in the number of employed in LFS 2001b (See Table 4.7) was mainly attributable to this occupation category.

Table 4.15: Percentage of employed in each broad occupation category, 1994-2009

	[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]	[K]
OHS 1994	5.5%	6.5%	8.7%	12.2%	10.4%	1.3%	12.3%	11.8%	20.1%	10.8%	0.5%
OHS 1995	5.3%	3.4%	11.1%	11.9%	11.4%	1.2%	11.8%	11.6%	24.7%	7.3%	0.2%
OHS 1996	4.9%	4.1%	13.7%	9.7%	11.6%	2.9%	13.0%	8.7%	16.8%	8.5%	6.0%
OHS 1997	7.3%	8.8%	8.3%	8.7%	10.3%	3.0%	14.4%	10.3%	16.6%	9.1%	3.1%
OHS 1998	7.7%	5.4%	9.6%	10.0%	12.3%	2.4%	14.0%	10.1%	17.8%	8.0%	2.6%
OHS 1999	6.6%	5.3%	10.0%	10.3%	11.8%	4.5%	13.1%	10.5%	18.2%	7.8%	1.7%
LFS 2000a	5.3%	3.7%	8.9%	8.8%	11.3%	14.0%	12.1%	9.5%	17.7%	8.4%	0.3%
LFS 2000b	4.7%	4.8%	9.3%	8.6%	12.0%	9.8%	13.0%	10.0%	19.7%	7.7%	0.5%
LFS 2001a	5.2%	3.8%	9.7%	8.7%	13.6%	7.7%	12.7%	9.5%	21.8%	6.9%	0.4%
LFS 2001b	5.9%	4.3%	10.5%	9.8%	12.8%	4.7%	13.7%	10.1%	20.1%	7.9%	0.2%
LFS 2002a	6.1%	4.1%	10.4%	9.5%	11.4%	9.1%	12.2%	10.0%	19.3%	7.5%	0.4%
LFS 2002b	6.5%	4.4%	10.7%	9.8%	11.0%	6.2%	12.9%	10.2%	20.3%	7.5%	0.4%
LFS 2003a	6.3%	4.9%	10.0%	9.7%	11.4%	3.8%	12.4%	10.6%	22.6%	7.8%	0.4%
LFS 2003b	7.2%	4.8%	10.1%	10.1%	11.9%	3.0%	12.7%	10.0%	22.1%	7.8%	0.2%
LFS 2004a	7.3%	4.7%	9.9%	10.3%	11.8%	2.7%	12.4%	10.2%	23.0%	7.4%	0.1%
LFS 2004b	7.8%	3.9%	9.9%	10.0%	12.5%	2.8%	13.2%	9.6%	22.5%	7.6%	0.2%
LFS 2005a	6.7%	4.5%	9.5%	10.1%	12.3%	3.6%	13.8%	9.8%	22.4%	7.1%	0.2%
LFS 2005b	7.0%	4.8%	9.7%	9.7%	13.1%	2.5%	14.2%	9.2%	22.8%	7.0%	0.2%
LFS 2006a	6.8%	4.8%	9.4%	9.7%	12.5%	5.2%	13.7%	8.8%	22.0%	6.8%	0.2%
LFS 2006b	6.8%	4.7%	9.6%	9.7%	12.8%	3.4%	15.0%	8.7%	22.2%	6.9%	0.1%
LFS 2007a	7.2%	4.7%	9.3%	10.0%	12.9%	3.3%	14.0%	9.2%	21.9%	7.4%	0.1%
LFS 2007b	7.6%	7.6%	10.1%	8.9%	12.1%	2.6%	13.8%	9.0%	20.2%	7.7%	0.3%
QLFS 2008Q1	7.0%	5.5%	10.5%	10.8%	13.2%	0.9%	14.3%	8.4%	22.4%	7.0%	0.0%
QLFS 2008Q2	7.2%	5.7%	10.6%	10.6%	12.7%	0.7%	14.2%	8.5%	22.7%	7.1%	0.0%
QLFS 2008Q3	7.7%	5.3%	10.9%	10.7%	13.0%	0.7%	13.8%	8.8%	21.6%	7.4%	0.0%
QLFS 2008Q4	7.8%	5.4%	11.0%	10.4%	12.5%	0.8%	13.6%	8.7%	22.2%	7.6%	0.0%
QLFS 2009Q1	7.8%	5.8%	11.0%	10.3%	13.0%	0.9%	13.3%	8.9%	21.4%	7.6%	0.0%
QLFS 2009Q2	7.7%	4.9%	11.6%	10.8%	13.5%	0.6%	13.3%	8.9%	21.6%	7.2%	0.0%
QLFS 2009Q3	7.7%	5.5%	11.6%	11.1%	13.9%	0.6%	12.2%	8.5%	21.8%	7.2%	0.0%
QLFS 2009Q4	7.6%	5.0%	11.7%	11.3%	14.2%	0.6%	12.1%	8.3%	22.2%	6.9%	0.0%

Highly-skilled occupations: [A]: Legislators, senior officials and managers; [B]: Professionals; [C]: Technicians and associate professionals.
Semi-skilled occupations: [D]: Clerks; [E]: Service workers and shop and market sales; [F]: Skilled agricultural and fishery worker; [G]: Craft and related trade workers; [H]: Plant and machinery operators and assemblers.
Unskilled occupations: [I]: Elementary occupations; [J]: Domestic workers.
Others: [K]: Others / Unspecified

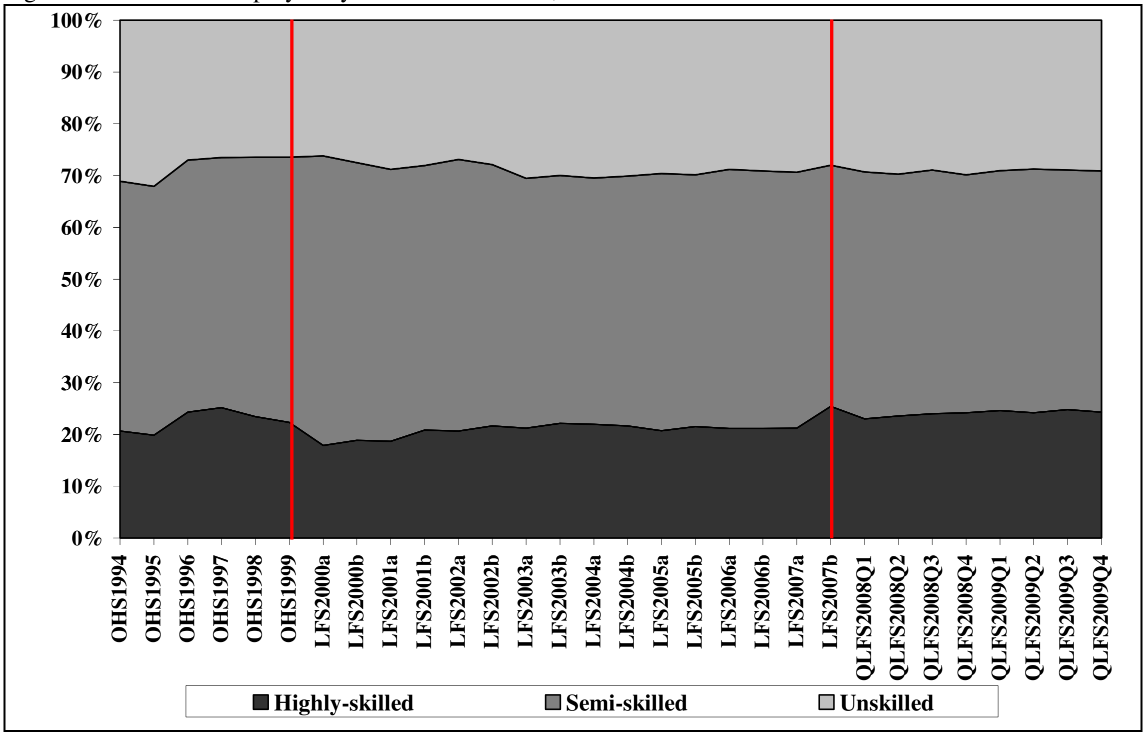
Looking at skills level of work, Table 4.16 and Figure 4.12 show that, although there was an increase in the number of people engaged in highly-skilled occupations throughout the years, skilled employment as percentage of total employment showed only a moderate increase of 3.6 percentage points between OHS 1994 and QLFS 2009Q4. Such increase could have been greater, as it is possible that unskilled employment as a share of overall employment was under-captured in the OHSs. This may well have resulted from the relatively poor capture of informal and low-income employment in these surveys (See Section 3.11).

Table 4.16: Employment by skills level of work, 1994-2009

	Unskilled (1 000s)	**Semi-skilled (1 000s)**	**Highly-Skilled (1 000s)**	**Highly-skilled as % of employed**
OHS 1994	2 922	4 529	1 942	20.7%
OHS 1995	3 045	4 553	1 884	19.9%
OHS 1996	2 274	4 107	2 044	24.3%
OHS 1997	2 339	4 251	2 220	25.2%
OHS 1998	2 414	4 574	2 136	23.4%
OHS 1999	2 696	5 206	2 277	22.4%
LFS 2000a	3 102	6 620	2 116	17.9%
LFS 2000b	3 347	6 526	2 292	18.8%
LFS 2001a	3 517	6 412	2 279	18.7%
LFS 2001b	3 130	5 691	2 323	20.8%
LFS 2002a	3 113	6 058	2 390	20.7%
LFS 2002b	3 136	5 669	2 437	21.7%
LFS 2003a	3 444	5 418	2 392	21.3%
LFS 2003b	3 421	5 451	2 522	22.1%
LFS 2004a	3 467	5 398	2 497	22.0%
LFS 2004b	3 496	5 596	2 515	21.7%
LFS 2005a	3 515	5 899	2 457	20.7%
LFS 2005b	3 666	5 962	2 639	21.5%
LFS 2006a	3 584	6 204	2 630	21.2%
LFS 2006b	3 722	6 349	2 703	21.2%
LFS 2007a	3 703	6 242	2 673	21.2%
LFS 2007b	3 712	6 171	3 370	25.4%
QLFS 2008Q1	4 000	6 498	3 138	23.0%
QLFS 2008Q2	4 095	6 410	3 242	23.6%
QLFS 2008Q3	3 959	6 433	3 275	24.0%
QLFS 2008Q4	4 137	6 369	3 354	24.2%
QLFS 2009Q1	3 968	6 328	3 356	24.6%
QLFS 2009Q2	3 855	6 298	3 235	24.2%
QLFS 2009Q3	3 736	5 967	3 194	24.8%
QLFS 2009Q4	3 786	6 042	3 156	24.3%

Table 4.16 also indicates that the rapid increase of employment between OHS 1999 and LFS 2000a was attributable to the increase of unskilled and semi-skilled employment. In contrast, the abrupt decrease of employment between LFS 2001a and LFS 2001b was caused by the decline of unskilled and semi-skilled employment. Finally, the rapid increase of highly-skilled employment in LFS 2007b accounted for the increase of total employment in this survey.

Figure 4.12: Share of employed by skills level of work, 1994-2009



A higher proportion of employed in the Indian and white population was engaged in highly-skilled occupations during the period under investigation (approximately 40% and 55% respectively in 2009, while such proportion was below 20% for both the black and coloured population), as these two race groups were more educated on average.

Table 4.17 shows the percentage of employed in each broad industry category, and it can be seen that the shares of transport / storage / communication as well as financial / insurance / business services increased during the period under study, at the cost of the agriculture / forestry / fishing / hunting and manufacturing industries.

The percentage of employed in the agriculture, forestry, fishing and hunting industry (i.e., column [A]) showed the greatest fluctuations. For example, this proportion decreased from 13.0% to 8.5% between OHS 1995 and OHS 1996, but increased from 10.6% in OHS 1999 to 19.2% in LFS 2000a. These abrupt changes in turn explain the rapid decline of employment between the 1995 and 1996 OHSs as well as the big increase of employment between OHS 1999 and LFS 2000a (See Table 2.7). This share dropped to 5.9% in QLFS 2008, and it was only during the QLFS years that this share stabilised at about 5%-6%.

Table 4.17: Percentage of employed in each broad industry category, 1994-2009

	[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]	[K]
OHS 1994	13.2%	2.8%	16.9%	1.0%	4.5%	17.4%	5.4%	6.1%	31.4%	0.3%	1.0%
OHS 1995	13.0%	4.6%	15.1%	0.9%	4.7%	17.5%	5.0%	6.1%	22.9%	8.4%	1.8%
OHS 1996	8.5%	2.8%	15.4%	1.4%	4.7%	15.3%	5.3%	8.3%	22.5%	9.0%	6.8%
OHS 1997	8.3%	4.3%	16.7%	1.3%	5.6%	17.3%	5.8%	8.0%	20.6%	8.3%	3.7%
OHS 1998	10.0%	4.6%	14.7%	1.2%	5.8%	19.0%	5.9%	9.1%	19.7%	8.2%	1.7%
OHS 1999	10.6%	4.6%	14.4%	0.8%	5.5%	20.0%	5.2%	9.0%	19.1%	9.3%	1.5%
LFS 2000a	19.2%	3.9%	12.4%	0.7%	5.0%	20.5%	4.6%	7.1%	16.0%	10.0%	0.6%
LFS 2000b	15.6%	4.9%	12.9%	0.8%	5.6%	20.2%	4.8%	8.0%	17.0%	9.4%	0.8%
LFS 2001a	12.8%	4.6%	13.2%	0.8%	5.2%	24.9%	4.7%	8.2%	16.4%	8.4%	0.6%
LFS 2001b	10.5%	5.0%	14.5%	0.8%	5.7%	22.0%	4.9%	9.3%	17.8%	9.2%	0.4%
LFS 2002a	15.0%	4.7%	13.8%	0.7%	5.0%	20.0%	4.9%	8.9%	17.3%	9.3%	0.5%
LFS 2002b	12.6%	5.0%	14.5%	0.7%	5.4%	19.4%	5.1%	9.6%	18.1%	9.1%	0.6%
LFS 2003a	11.4%	4.9%	14.0%	0.8%	5.2%	20.6%	5.1%	9.2%	18.7%	9.6%	0.4%
LFS 2003b	10.6%	4.8%	13.6%	0.8%	5.8%	21.3%	4.7%	9.6%	19.1%	9.4%	0.3%
LFS 2004a	11.0%	4.9%	14.0%	0.9%	5.8%	20.7%	5.1%	9.4%	18.9%	9.0%	0.2%
LFS 2004b	9.1%	3.5%	14.7%	0.9%	7.1%	21.8%	4.8%	9.9%	18.8%	9.2%	0.2%
LFS 2005a	9.8%	3.6%	13.9%	1.0%	6.8%	22.2%	5.0%	9.6%	18.8%	9.0%	0.2%
LFS 2005b	7.5%	3.3%	13.9%	0.8%	7.6%	24.6%	5.0%	10.5%	17.8%	8.7%	0.2%
LFS 2006a	10.6%	3.2%	13.9%	0.8%	6.9%	24.1%	4.5%	9.6%	17.5%	8.7%	0.2%
LFS 2006b	8.5%	3.1%	13.6%	0.9%	8.0%	23.9%	4.8%	10.2%	18.1%	8.7%	0.3%
LFS 2007a	8.5%	3.6%	13.9%	0.8%	7.6%	23.4%	4.6%	10.4%	18.3%	8.8%	0.1%
LFS 2007b	7.8%	3.2%	13.2%	0.7%	7.9%	22.1%	5.2%	11.1%	19.2%	9.0%	0.4%
QLFS 2008Q1	5.9%	2.4%	14.6%	0.7%	8.2%	23.2%	5.5%	12.2%	18.8%	8.5%	0.0%
QLFS 2008Q2	5.8%	2.5%	14.3%	0.7%	8.3%	22.6%	5.7%	12.3%	19.2%	8.6%	0.0%
QLFS 2008Q3	5.6%	2.3%	14.0%	0.7%	8.1%	23.3%	5.6%	12.0%	19.1%	9.3%	0.0%
QLFS 2008Q4	5.5%	2.3%	14.0%	0.6%	8.6%	22.9%	5.6%	11.8%	19.2%	9.4%	0.0%
QLFS 2009Q1	5.4%	2.4%	13.8%	0.7%	8.3%	22.1%	5.6%	12.7%	19.5%	9.5%	0.0%
QLFS 2009Q2	5.3%	2.4%	14.0%	0.7%	8.3%	22.2%	5.4%	12.8%	20.0%	8.9%	0.0%
QLFS 2009Q3	5.1%	2.3%	13.4%	0.6%	8.2%	22.1%	5.7%	13.1%	20.4%	9.0%	0.1%
QLFS 2009Q4	4.7%	2.3%	13.4%	0.8%	8.4%	22.1%	5.7%	13.6%	20.3%	8.7%	0.0%

Primary sector:	[A]: Agriculture, forestry, fishing and hunting
	[B]: Mining and quarrying
Secondary sector:	[C]: Manufacturing
	[D]: Electricity, gas and water supply
	[E]: Construction
Tertiary sector:	[F]: Wholesale and retail
	[G]: Transport, storage and communication
	[H]: Financial, insurance and business services
	[I]: Community, social and personal services
Other:	[J]: Private households
	[K]: Other / Unspecified

The changing nature of employment by the three broad skills categories at the industry level in selected years is presented in Table 4.18 and Table A.10 in Appendix A. As mentioned before, there was a slight increase in the proportion of highly-skilled employed of 3.6 percentage points if only OHS 1995 and QLFS 2009Q4 are compared, at the expense of the unskilled employed (decreasing by 1.9 percentage points to 29.2% in QLFS 2009Q4) and the semi-skilled employed (decreasing by 1.7 percentage points to 46.5% in QLFS 2009Q4).

Table 4.18: Skills breakdown of employment by industry, selected surveys

Industry	Skills	OHS 1995	OHS 1997	OHS 1999	LFS 2001b	LFS 2003b	LFS 2005b	LFS 2007b	QLFS 2009Q4
[A]	Highly-skilled	0.8%	5.0%	3.3%	1.4%	5.2%	4.9%	3.5%	5.9%
	Semi-skilled	22.0%	40.6%	42.9%	40.9%	41.3%	44.0%	44.2%	20.9%
	Unskilled	77.2%	54.4%	53.8%	57.7%	53.5%	51.2%	52.3%	73.1%
[B]	Highly-skilled	6.7%	18.7%	8.6%	5.9%	7.3%	6.1%	7.9%	11.3%
	Semi-skilled	74.2%	64.6%	82.1%	83.6%	81.4%	78.5%	74.6%	71.9%
	Unskilled	19.2%	16.7%	9.2%	10.5%	11.3%	15.5%	17.5%	16.8%
[C]	Highly-skilled	11.7%	18.4%	17.2%	16.4%	18.3%	15.7%	19.2%	19.5%
	Semi-skilled	68.5%	59.4%	65.3%	67.3%	66.7%	66.0%	64.7%	64.9%
	Unskilled	19.8%	22.2%	17.4%	16.3%	15.0%	18.3%	16.1%	15.7%
[D]	Highly-skilled	18.1%	19.1%	19.2%	22.5%	21.0%	22.6%	32.0%	27.9%
	Semi-skilled	67.0%	62.4%	62.9%	64.6%	65.8%	63.3%	57.0%	58.1%
	Unskilled	14.9%	18.5%	18.0%	12.9%	13.2%	14.1%	10.9%	14.0%
[E]	Highly-skilled	9.5%	8.7%	9.2%	7.2%	8.2%	9.6%	10.8%	12.2%
	Semi-skilled	70.9%	72.8%	73.0%	77.1%	68.5%	67.1%	62.4%	60.7%
	Unskilled	19.6%	18.5%	17.8%	15.7%	23.3%	23.3%	26.8%	27.1%
[F]	Highly-skilled	16.7%	19.2%	14.6%	12.4%	13.0%	13.5%	17.8%	15.1%
	Semi-skilled	63.5%	57.5%	59.4%	56.6%	55.0%	53.3%	54.1%	55.1%
	Unskilled	19.8%	23.3%	26.0%	31.0%	32.0%	33.3%	28.1%	29.8%
[G]	Highly-skilled	26.2%	21.5%	23.5%	25.7%	24.3%	20.9%	30.3%	25.4%
	Semi-skilled	61.9%	66.8%	67.2%	63.6%	62.7%	62.8%	59.5%	58.1%
	Unskilled	12.0%	11.7%	9.3%	10.7%	13.0%	16.2%	10.2%	16.5%
[H]	Highly-skilled	37.6%	39.4%	42.6%	43.3%	44.6%	41.7%	44.8%	39.0%
	Semi-skilled	55.8%	49.2%	49.3%	47.6%	45.1%	48.1%	45.8%	45.8%
	Unskilled	6.6%	11.4%	8.1%	9.1%	10.3%	10.3%	9.4%	15.2%
[I]	Highly-skilled	45.6%	54.1%	53.0%	52.5%	51.4%	50.8%	55.5%	48.4%
	Semi-skilled	39.2%	31.6%	35.9%	36.3%	35.7%	36.0%	33.4%	38.8%
	Unskilled	15.2%	14.4%	11.1%	11.2%	12.9%	13.2%	11.2%	12.8%
[J]	Highly-skilled	0.2%	1.3%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
	Semi-skilled	2.4%	16.0%	16.2%	14.7%	0.5%	0.8%	0.4%	1.2%
	Unskilled	97.5%	82.7%	83.7%	85.3%	99.5%	99.2%	99.6%	98.8%
All employed	Highly-skilled	19.9%	25.2%	22.4%	20.9%	22.1%	21.5%	25.4%	24.3%
	Semi-skilled	48.0%	48.3%	51.2%	51.1%	47.8%	48.6%	46.6%	46.5%
	Unskilled	32.1%	26.6%	26.5%	28.1%	30.0%	29.9%	28.0%	29.2%

Primary sector: [A]: Agriculture, forestry, fishing and hunting
[B]: Mining and quarrying
Secondary sector: [C]: Manufacturing
[D]: Electricity, gas and water supply
[E]: Construction
Tertiary sector: [F]: Wholesale and retail
[G]: Transport, storage and communication
[H]: Financial, insurance and business services
[I]: Community, social and personal services
Other: [J]: Private households

However, the experiences were varied when looking at the skills composition of each industry. In the primary sector, only a small proportion (less than 6% in general) of workers in the agricultural, forest, fishing and hunting industry is involved in highly-skilled occupations. In mining and quarrying, employment shifted slightly in favour of semi-skilled occupations against the unskilled occupations during the OHSs and the early LFSs, but if one looks at LFS 2007b and QLFS 2009Q4, the proportions are largely similar to those in OHS 1995. In the secondary sector, in manufacturing as well as electricity, gas and water supply, the highly-skilled proportion of the employed increased slightly. In construction, the proportion of highly-skilled employment started to increase (exceeding 10%) since LFS 2007b.

Looking at the tertiary sector, the highly-skilled share of the employed in the wholesale and retail industry and the transport, storage and communication industry have been hovering around the 10%-15% and 25%-30% ranges during the period under study. On the other hand, the financial, insurance and business services as well as the community, social and personal services industries are the two categories with the highest proportion of highly-skilled workers (approximately 40% and 50% respectively).

4.5 Characteristics of the unemployed

This section examines the levels and trends of unemployment by using OHS 1995-QLFS 2009Q4 data, as well as the demographic, location and educational attainment characteristics of the unemployed. The person weight variable is used to derive the weighted figures.

4.5.1 Number of unemployed and unemployment rates

Table 4.19 and Figure 4.13 show that the number of narrowly defined unemployed more than doubled from 2 million in OHS 1995 to above 4 million in QLFS 2009Q4, while the number of broadly defined unemployed also increased from 4.2 million to 7.3 million between the OHS 1995 and LFS 2007b, before an abrupt decrease of nearly 2 million was recorded during the changeover from LFS to QLFS. This was caused by the different methodology adopted in QLFS to capture the labour market status, as mentioned in Section 3.11.

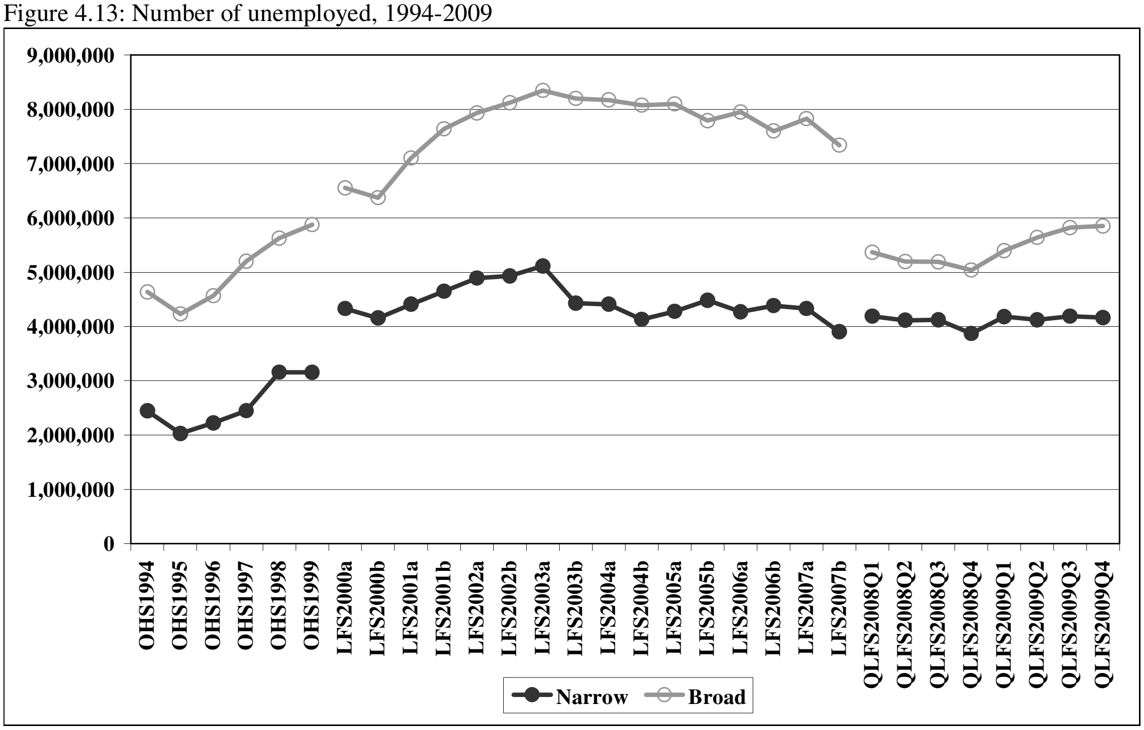
Table 4.19: Number of unemployed (1 000s), 1994-2009

	Number (1 000s)		Absolute change (1 000s)		Percentage change	
	Narrow	Broad	Narrow	Broad	Narrow	Broad
OHS 1994	2 445	4 634				
OHS 1995	2 028	4 232	-417	-402	-17.0%	-8.7%
OHS 1996	2 224	4 566	196	335	9.7%	7.9%
OHS 1997	2 451	5 202	226	636	10.2%	13.9%
OHS 1998	3 158	5 626	707	425	28.9%	8.2%
OHS 1999	3 154	5 875	-4	249	-0.1%	4.4%
LFS 2000a	4 331	6 550	1 177	675	37.3%	11.5%
LFS 2000b	4 157	6 372	-174	-178	-4.0%	-2.7%
LFS 2001a	4 408	7 101	251	729	6.0%	11.4%
LFS 2001b	4 650	7 640	242	539	5.5%	7.6%
LFS 2002a	4 891	7 932	241	292	5.2%	3.8%
LFS 2002b	4 931	8 121	40	189	0.8%	2.4%

Table 4.19: Continued

	Number (1 000s)		Absolute change (1 000s)		Percentage change	
	Narrow	Broad	Narrow	Broad	Narrow	Broad
LFS 2003a	5 111	8 345	181	224	3.7%	2.8%
LFS 2003b	4 429	8 198	-682	-146	-13.3%	-1.8%
LFS 2004a	4 410	8 172	-20	-27	-0.4%	-0.3%
LFS 2004b	4 131	8 074	-279	-97	-6.3%	-1.2%
LFS 2005a	4 278	8 098	147	24	3.6%	0.3%
LFS 2005b	4 482	7 791	204	-307	4.8%	-3.8%
LFS 2006a	4 270	7 949	-212	158	-4.7%	2.0%
LFS 2006b	4 386	7 599	116	-350	2.7%	-4.4%
LFS 2007a	4 331	7 830	-55	231	-1.3%	3.0%
LFS 2007b	3 901	7 340	-430	-491	-9.9%	-6.3%
QLFS 2008Q1	4 189	5 366	288	-1 974	7.4%	-26.9%
QLFS 2008Q2	4 115	5 193	-75	-173	-1.8%	-3.2%
QLFS 2008Q3	4 120	5 191	6	-1	0.1%	0.0%
QLFS 2008Q4	3 871	5 039	-249	-152	-6.0%	-2.9%
QLFS 2009Q1	4 181	5 396	310	356	8.0%	7.1%
QLFS 2009Q2	4 123	5 639	-58	243	-1.4%	4.5%
QLFS 2009Q3	4 190	5 820	67	181	1.6%	3.2%
QLFS 2009Q4	4 162	5 847	-28	27	-0.7%	0.5%

Figure 4.13: Number of unemployed, 1994-2009



Throughout the years, the number of unemployed was extremely unstable. For example, the increase of the number of unemployed was relatively more rapid between OHS 1995 and LFS 2000a in both narrow and broad terms than it was in the surveys following LFS 2000a. After an unusual decrease in LFS 2000b, these figures displayed an increasing trend again until LFS 2003a.

With regard to the possible causes of the general upward trend of unemployment between OHS 1995 and LFS 2003a, as in the discussion on LF and employment trends, it could be attributed to the changes in labour market status derivation methodology, adoption of different sampling techniques across the surveys, or it was real as a result of what happened to the economy at the time. Looking at the last reason in greater detail, as explained in footnote 64, the policies implemented by the South African Department of Education in the late 1990s led to the over-aged students being pushed out of the school system. This possibly led to these youngsters entering the labour market for work, yet most of them had low employment prospects, thereby causing the unemployment to up in the last few OHSs and the initial LFSs.

Since LFS 2003b, the number of narrow unemployed seemed to have stabilised at between 4.2-4.4 million in the last few LFSs, while there was a slight downward trend in the number of broad unemployed from slightly above 8 million in 2002-2005 to 7.34 million in LFS 2007b. Finally, with the introduction of QLFS, the number of narrow unemployed remained approximately 4.2 million in 2008-2009, while the number of broad unemployed increased rapidly from 5.37 million in QLFS 2008Q1 to 5.85 million in QLFS 2009Q4, as a result of the impact of global recession.

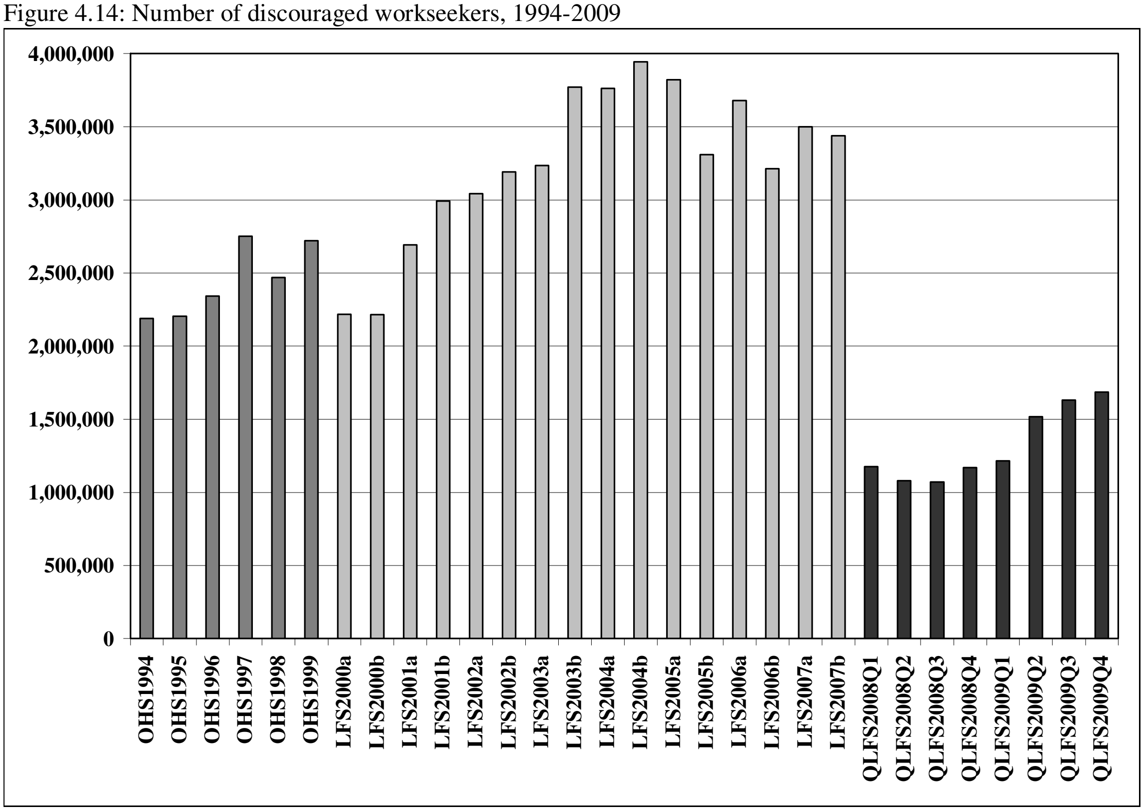
Table 4.20 and Figure 4.14 present the number of discouraged workseekers (the difference between the broad unemployed and narrow unemployed) in each survey. An upward trend took place in general between 1994 and 2004, peaking at 3.94 million in LFS 2004b, before a slight downward trend took place until LFS 2007b. As explained in Section 3.11, due to the adoption of a different and stricter labour market status derivation methodology since the launch of QLFS, the number of discouraged workseekers suddenly dropped by 65.78% during the changeover from LFS to QLFS, decreasing from 3.44 million in LFS 2007b to 1.18 million in QLFS 2008Q1. It fluctuated around the 1.07-1.22 million range until the first QLFS of 2009, before showing a continuous upward trend, increasing to 1.69 million in QLFS 2009Q4. This could be due to the fact that some labour force participants felt pessimistic that they could find employment as a result of the 2008 recession, thereby stopping to seek work actively.

Section 4.6 will examine whether the abrupt break in the number of discouraged workseekers between LFS 2007b and QLFS 2008Q1 would not have taken place, had a consistent labour market status derivation methodology been applied across all surveys, if possible.

Table 4.20: Number of discouraged workseekers (1 000s), 1994-2009

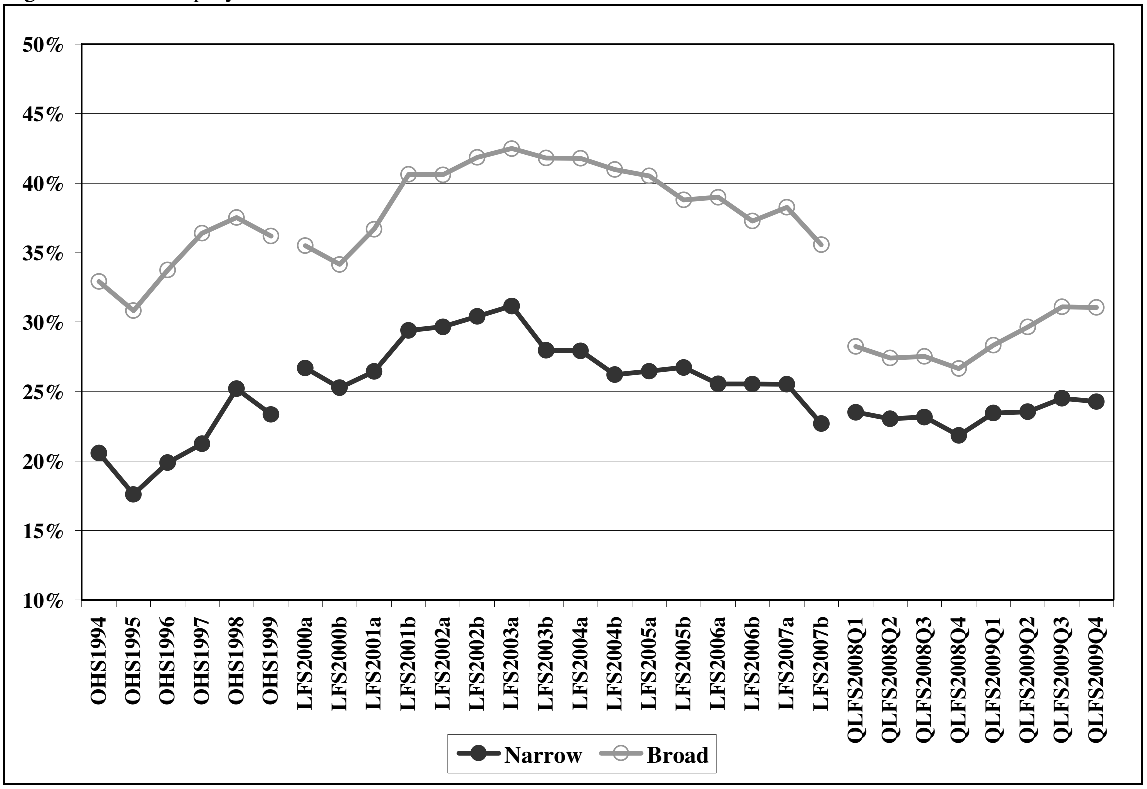
	Number (1 000s)	Absolute change (1 000s)	Percentage change		Number (1 000s)	Absolute change (1 000s)	Percentage change
OHS1994	2189			LFS2004b	3943	181	4.82%
OHS1995	2203	14	0.65%	LFS2005a	3819	-124	-3.14%
OHS1996	2342	139	6.29%	LFS2005b	3308	-511	-13.38%
OHS1997	2751	409	17.47%	LFS2006a	3679	371	11.20%
OHS1998	2469	-283	-10.28%	LFS2006b	3213	-466	-12.67%
OHS1999	2721	253	10.24%	LFS2007a	3499	286	8.90%
LFS2000a	2218	-503	-18.48%	LFS2007b	3439	-60	-1.73%
LFS2000b	2215	-4	-0.16%	QLFS2008Q1	1177	-2262	-65.78%
LFS2001a	2693	478	21.59%	QLFS2008Q2	1078	-99	-8.39%
LFS2001b	2991	297	11.04%	QLFS2008Q3	1071	-7	-0.65%
LFS2002a	3041	51	1.69%	QLFS2008Q4	1168	97	9.09%
LFS2002b	3190	149	4.90%	QLFS2009Q1	1215	47	3.99%
LFS2003a	3233	43	1.35%	QLFS2009Q2	1516	301	24.79%
LFS2003b	3769	536	16.57%	QLFS2009Q3	1631	114	7.53%
LFS2004a	3762	-7	-0.19%	QLFS2009Q4	1685	54	3.34%

Figure 4.14: Number of discouraged workseekers, 1994-2009



Looking at the trends in unemployment rates, Figure 4.15 shows that, despite the fluctuations explained above, both the narrow and broad unemployment rates displayed an upward trend before peaking in LFS 2003a, i.e., jobless growth under Altman's second definition (real GDP growth complemented by a rising unemployment rate) took place. From then onwards, both rates displayed a continuous downward trend. In LFS 2007b, the narrow and broad unemployment rates were 22.7% and 35.6% respectively. Looking at the narrow unemployment rates since the inception of the QLFS, a very slight upward trend is observed in 2009, which could be attributed to the impact of the global economic recession. Although the post-2007 QLFS broad unemployment rates are not comparable with the OHS/LFS broad unemployment rates, it also displayed an upward trend in 2009 (an increase of more than five percentage points).

Figure 4.15: Unemployment rates, 1994-2009



4.5.2 Demographic, geographic and educational attainment characteristics of the unemployed

The trends in both the narrow and broad unemployment rates discussed in Section 3.4.1 were also observed for both genders between 1994 and 2009 (Figure 4.16). With regard to the gender share of unemployed, the female share stabilised at about 50% during the OHSs, and increased slightly to about 55% in the LFSs, before dropping to slightly above 50% in the QLFSs. Thus, these trends once again do not suggest a feminisation of the labour force or employment, or that the female share of the unemployed would have shown an obvious decline.

Figure 4.16: Unemployment rates by gender, 1994-2009

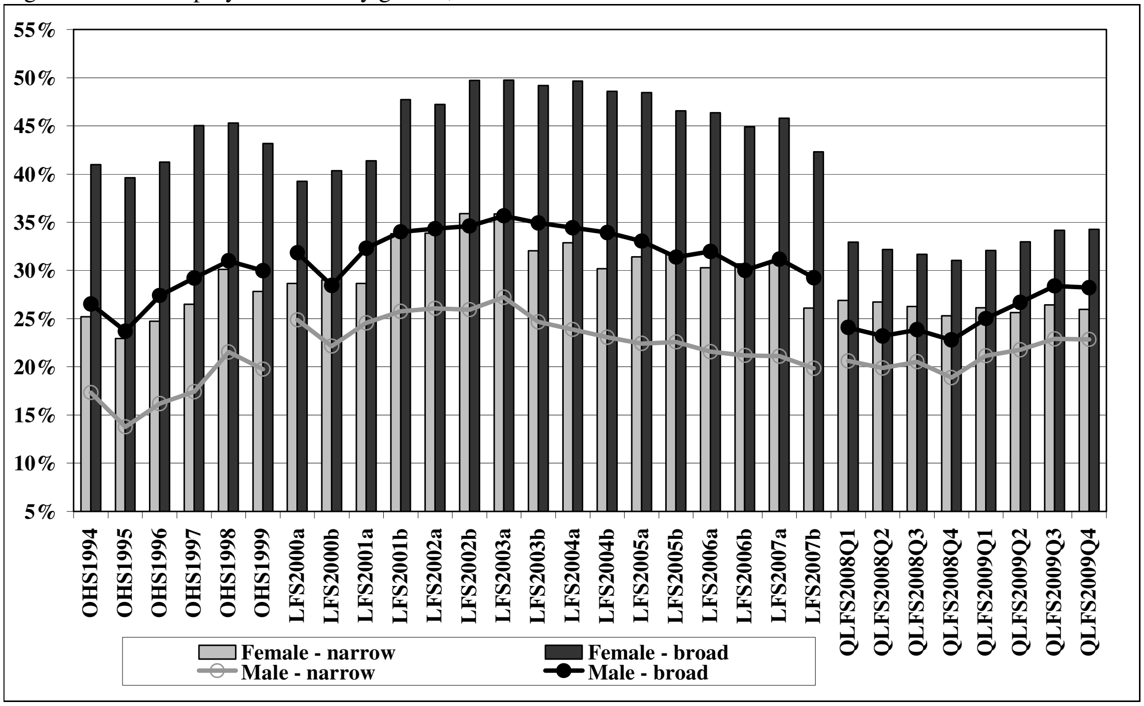
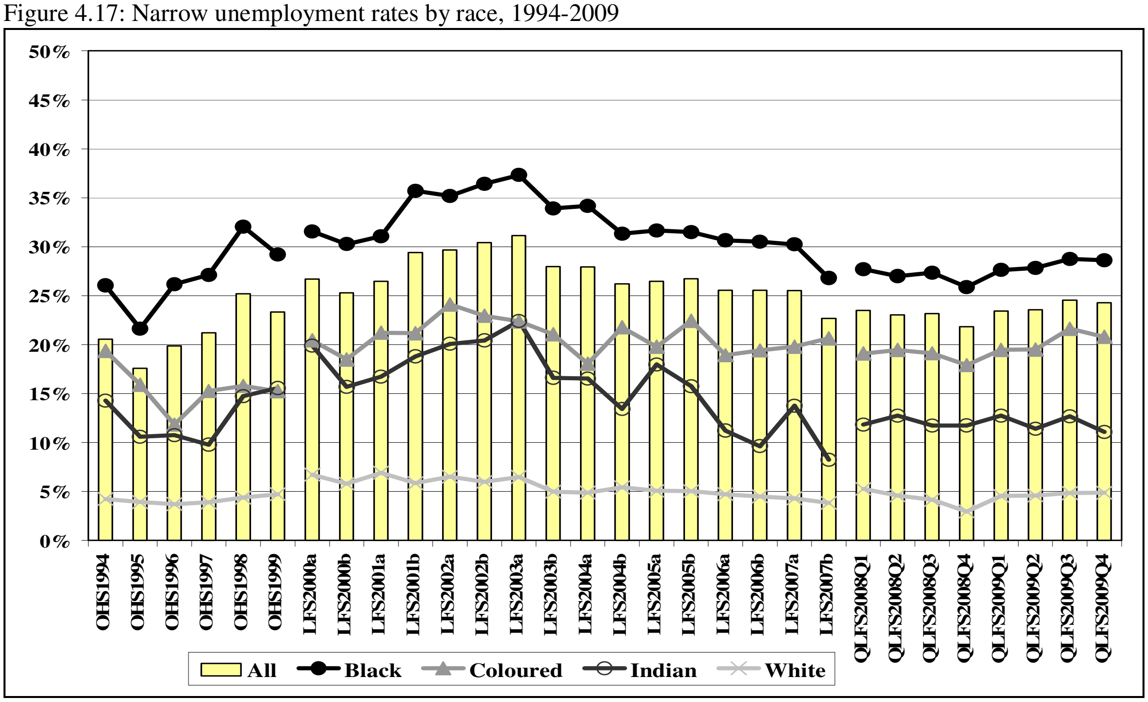


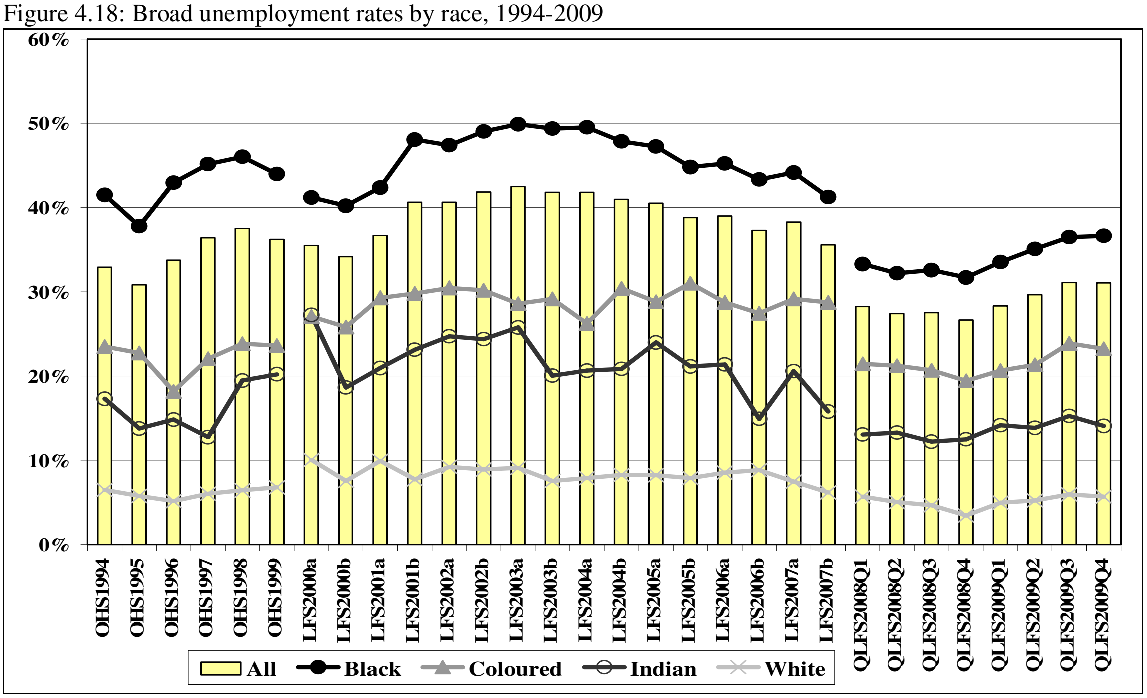
Figure 4.17 shows that, in narrow terms, the white unemployment rate has been stagnant at approximately 5% during the 1994-2009 period, while the coloured and Indian rates increased until LFS 2003a, before stabilising at about 20% and 12% respectively. Interestingly, the black unemployment rate also showed a similar upward trend until peaking in LFS 2003a; a downward trend took place before it stabilised in the 25%-30% ranges since LFS 2007b.

Figure 4.17: Narrow unemployment rates by race, 1994-2009



Figures 4.18 shows that, in broad terms, the highest unemployment rates (in excess of 40% in most surveys) were experienced by blacks. However, these have shown a slight declining trend since LFS 2004b. The coloured unemployment rate showed an upward trend until LFS 2005b, while the Indian unemployment rate was extremely unstable (this could be due to the small Indian sample size in the surveys under study). The black share of the unemployed remained quite stable at slightly below 90% of the total unemployed in both narrow and broad terms throughout the years in question.

Figure 4.18: Broad unemployment rates by race, 1994-2009



Narrow unemployment rates by province are presented in Table 4.21. It can be seen that these rates in Northern Cape, Free State, North West and Limpopo experienced the greatest increase since 1994. Western Cape and Gauteng were the two provinces with the lowest unemployment rates. About 60% of the unemployed were concentrated in the following three provinces: Western Cape, KwaZulu-Natal and Gauteng.

As discussed earlier (See Figure 4.15), the narrow unemployment increased abruptly between OHS 1999 and LFS 2000a. This could be attributable to the increase of unemployment rates in all provinces except Eastern Cape and Limpopo (the unemployment rate of these two provinces experienced an unusual abrupt decrease) as well as Free State, as shown in Table 4.21.

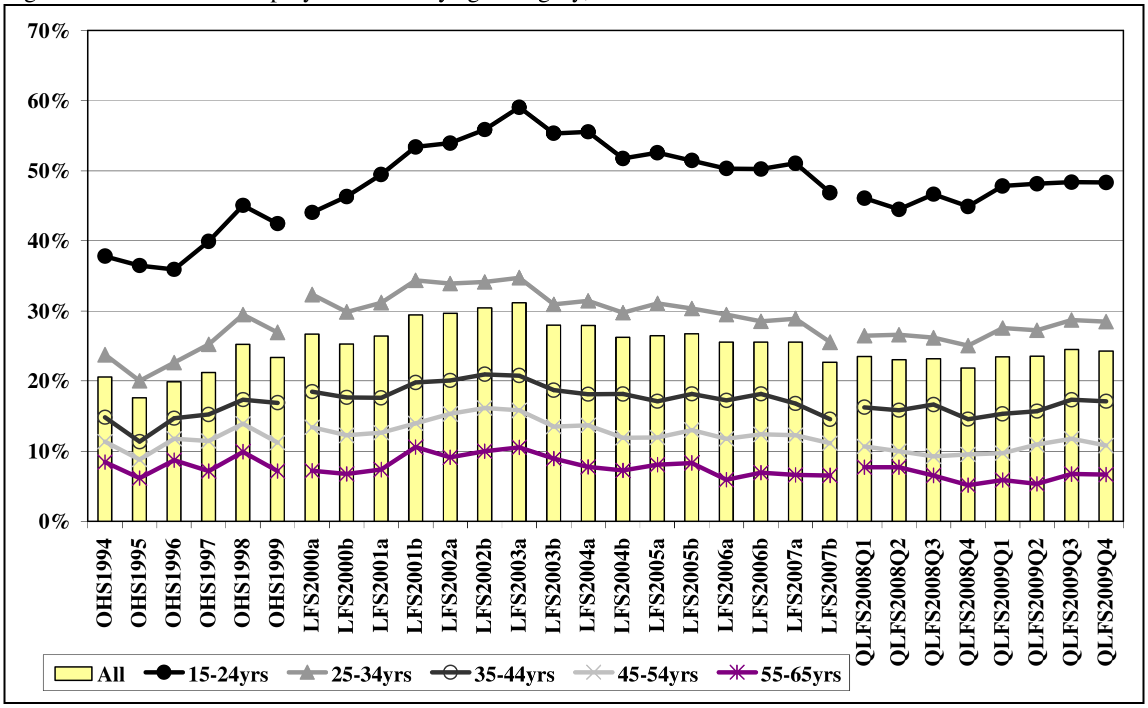
Table 4.21: Narrow unemployment rates by province, 1994-2009

	WC	EC	NC	FS	KZN	NW	GAU	MPU	LIM	RSA
OHS 1994	12.9%	28.5%	28.5%	13.9%	19.8%	23.7%	18.3%	25.7%	28.2%	20.6%
OHS 1995	13.8%	24.3%	20.4%	12.4%	20.7%	17.2%	15.7%	16.5%	20.0%	17.6%
OHS 1996	11.2%	28.5%	12.2%	20.0%	24.2%	15.9%	19.1%	14.1%	29.1%	19.9%
OHS 1997	11.8%	28.8%	18.4%	18.9%	22.9%	21.3%	21.4%	23.4%	25.8%	21.2%
OHS 1998	13.5%	36.9%	17.9%	21.3%	27.2%	26.5%	23.1%	25.0%	35.7%	25.2%
OHS 1999	13.7%	29.8%	18.1%	23.3%	25.9%	23.5%	20.6%	24.4%	34.0%	23.3%
LFS 2000a	18.9%	23.8%	23.1%	22.8%	30.8%	31.3%	28.9%	29.1%	25.0%	26.7%
LFS 2000b	16.5%	25.3%	18.5%	23.5%	27.7%	28.5%	26.5%	26.5%	27.9%	25.3%
LFS 2001a	19.0%	28.4%	23.8%	27.4%	26.2%	27.9%	28.2%	26.3%	28.1%	26.4%
LFS 2001b	17.7%	31.4%	25.0%	27.0%	33.8%	28.6%	30.4%	29.2%	34.6%	29.4%
LFS 2002a	18.4%	26.4%	27.7%	31.1%	35.9%	30.3%	29.8%	29.9%	35.1%	29.7%
LFS 2002b	19.6%	32.7%	24.9%	29.1%	35.0%	30.6%	30.5%	30.1%	34.1%	30.4%
LFS 2003a	19.9%	29.8%	28.9%	31.2%	36.3%	32.3%	30.8%	30.8%	39.4%	31.1%
LFS 2003b	19.5%	31.7%	26.4%	28.0%	31.6%	28.4%	27.6%	25.3%	31.1%	28.0%
LFS 2004a	16.8%	32.6%	22.4%	26.0%	33.0%	30.8%	27.7%	25.7%	31.2%	27.9%
LFS 2004b	18.6%	29.6%	24.5%	28.6%	28.7%	28.0%	25.7%	24.8%	27.8%	26.2%
LFS 2005a	17.6%	27.1%	29.4%	30.6%	31.7%	28.8%	22.7%	27.4%	32.4%	26.5%
LFS 2005b	18.9%	29.9%	24.7%	30.3%	32.8%	27.4%	22.8%	26.9%	30.1%	26.7%
LFS 2006a	15.9%	22.1%	23.5%	28.3%	29.9%	31.8%	23.3%	27.4%	35.6%	25.6%
LFS 2006b	15.0%	32.0%	28.7%	26.5%	26.6%	29.7%	23.2%	28.0%	32.0%	25.5%
LFS 2007a	17.2%	25.5%	26.5%	26.4%	29.2%	32.0%	22.6%	26.3%	32.4%	25.5%
LFS 2007b	15.7%	26.1%	26.0%	25.2%	30.0%	24.6%	17.4%	22.0%	27.3%	22.7%
QLFS 2008Q1	18.0%	28.1%	24.9%	25.0%	22.7%	22.2%	22.6%	23.6%	31.7%	23.5%
QLFS 2008Q2	19.1%	24.8%	24.7%	25.9%	22.2%	22.9%	21.8%	24.7%	30.5%	23.0%
QLFS 2008Q3	19.6%	27.4%	22.7%	22.9%	22.0%	26.7%	21.8%	23.1%	29.5%	23.2%
QLFS 2008Q4	16.9%	25.2%	21.5%	22.5%	20.7%	25.7%	20.6%	23.0%	28.8%	21.8%
QLFS 2009Q1	18.3%	28.4%	27.3%	25.4%	22.6%	26.8%	21.7%	24.7%	28.1%	23.4%
QLFS 2009Q2	20.5%	27.9%	26.4%	26.9%	19.3%	27.6%	23.1%	26.5%	24.8%	23.5%
QLFS 2009Q3	22.5%	26.8%	30.0%	28.6%	18.7%	27.9%	25.8%	25.6%	25.5%	24.5%
QLFS 2009Q4	21.4%	27.0%	24.8%	25.3%	19.2%	26.9%	25.7%	26.5%	26.9%	24.3%

WC: Western Cape
EC: Eastern Cape
NC: Northern Cape
FS: Free State
KZN: KwaZulu-Natal
NW: North West
GAU: Gauteng
MPU: Mpumalanga
LIM: Limpopo
RSA: South Africa

Unemployment rates did not show any fluctuations in the older age groups, as shown in Figure 4.19. These rates stabilised at approximately 18%, 12% and 8% in the 35-44 years, 45-54 years and 55-65 year old age categories respectively. On the other hand, the upward trend of the unemployment rate until LFS 2003a as discussed in Section 3.4.1 was relatively greater in the 15-24 year old age group. Consequently, the share of the unemployed accounted for by this age group increased slightly during the period under study (from 28% in OHSs to about 33% in QLFSs).

Figure 4.19: Narrow unemployment rates by age category, 1994-2009



It is expected that as the South African economy becomes more skill-intensive, the unemployment problem would become more serious for less educated people. Surprisingly, however, Figure 4.20 and Table A.11 in Appendix A show that even people with post-Matric qualifications initially showed an upward trend in unemployment, before a downward trend took place since LFS 2003b[71]. Furthermore, the share of unemployed with at least Matric increased from one-fifth in OHS 1994 to nearly 40% in QLFS 2009Q4 (Figure 4.21). Figure A.9 in Appendix A provides more information by presenting the proportion of unemployed with at least Matric by race throughout the years, and it can be seen that this proportion in the blacks almost doubled to 40% in 2009, while this proportion hovers around 60%-70% in most the surveys under study in the case of white unemployed.

[71] Pauw et al. (2006) identify a number of factors accounting for increasing graduate unemployment, such as the oversupply of graduates in certain fields of study (e.g., commerce), continued discrimination favouring Whites, lack of soft skills (e.g., communication skills, presentation skills, time management skills, basic numeracy and literacy skills, etc.), graduate over-expectation, etc. A recent report by the Centre for Development and Enterprise (2007) claims that the problem in the South African labour market is not only skills shortage (numbers of qualified and experienced people) but a skills deficit (poor skills quality of educated people), resulting in unemployment of seemingly ‘qualified’ people at both school-leaving and tertiary level.

Figure 4.20: Narrow unemployment rates by educational attainment, 1994-2009

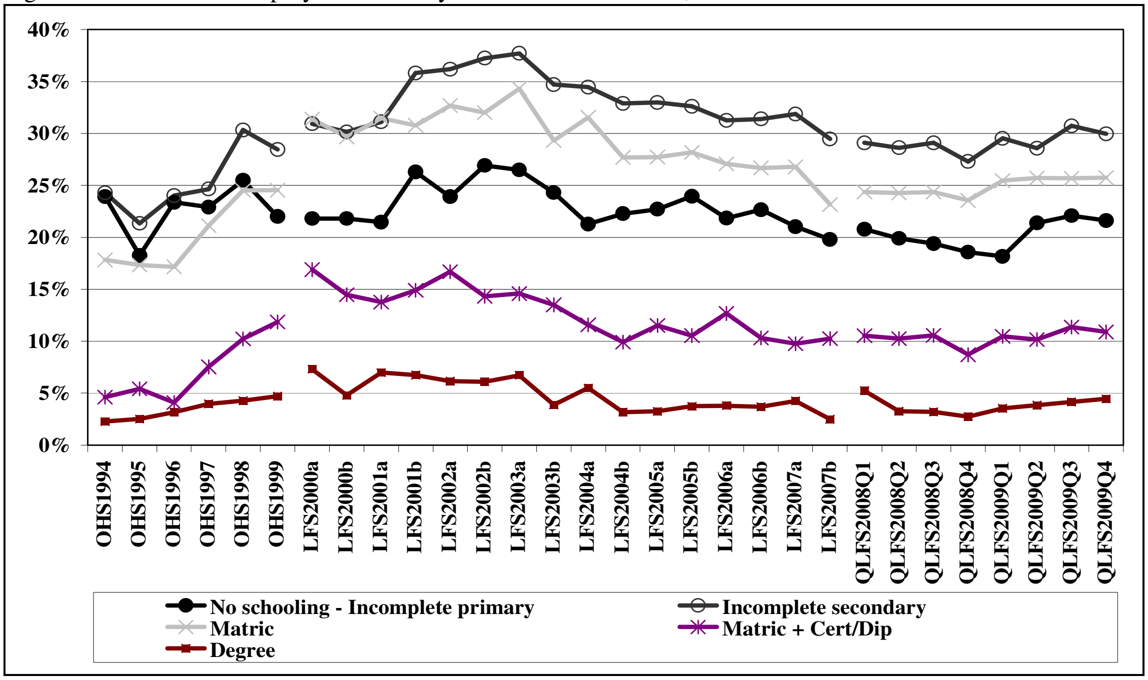
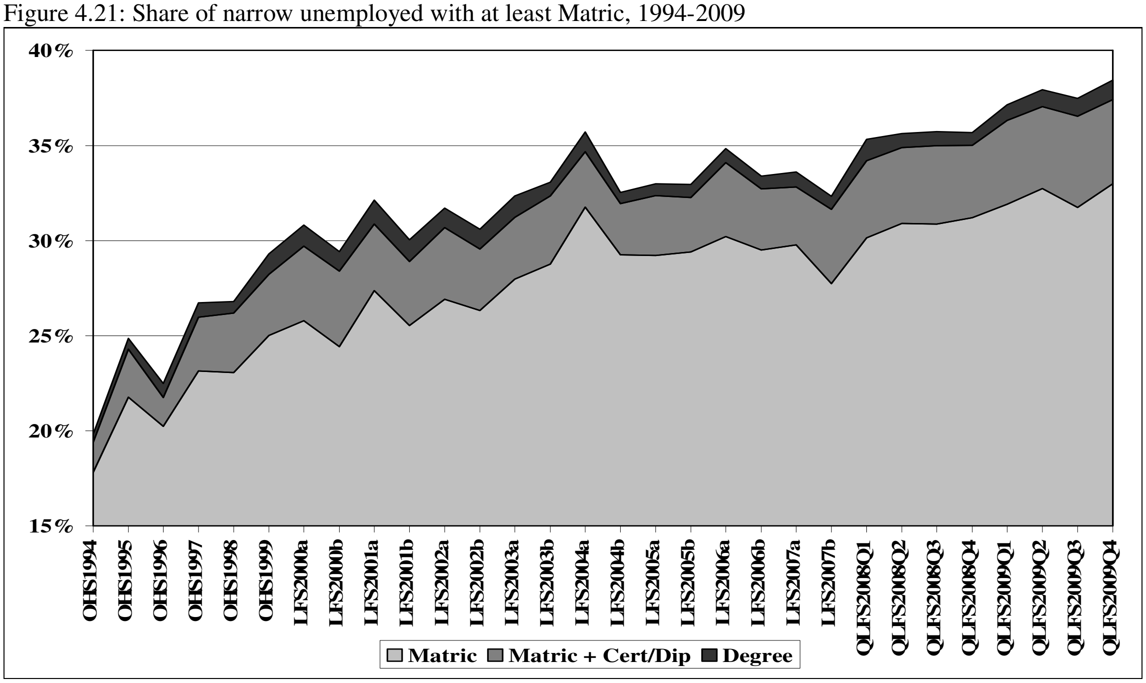


Figure 4.21: Share of narrow unemployed with at least Matric, 1994-2009



To conclude, both the narrow and broad unemployment rates have displayed an upward trend before peaking in LFS 2003a. From then onwards, both rates displayed a continuous downward trend. However, both rates increased again in 2009. The upward trends of both narrow and broad unemployment rates could make it more difficult for the ASGISA goal of reducing the narrow unemployment rate to below 15% by 2014 to be achieved. Furthermore, even if the goal of the New Growth Path, namely the creation of five million jobs by the end of 2021, is successfully met, the number of unemployed as well as unemployment rate might still be high enough to have negative impact on poverty and inequality significantly, and more than five million jobs might actually need to be created to alleviate the situation. Finally, after examining all available labour surveys, it was found that female blacks aged below 35 years at the time of the survey, without post-Matric qualifications and residing in provinces other than Western Cape and Gauteng are associated with a greater likelihood of unemployment.

4.6 Labour market trends revisited

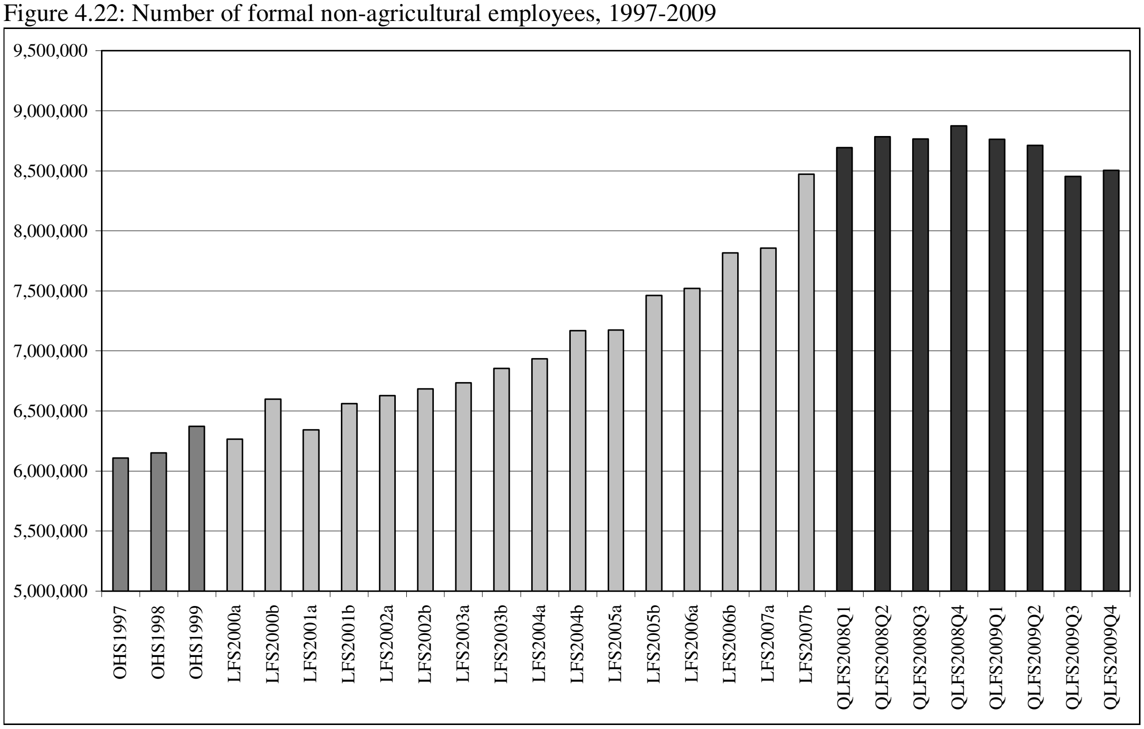
The results of the descriptive analyses on the labour market trends by using all available OHSs, LFSs and QLFSs in Sections 4.3-4.5 found that there was an abrupt increase between OHS 1997 and LFS 2000a in LF and LFPRs in both narrow and broad terms, and such increase was most rapid during the changeover between OHS and LFS. In addition, the number of unemployed as well as unemployment rates in both narrow and broad terms increased quite rapidly between OHS 1995 and OHS 1998, and the increase was very rapid between OHS 1999 and LFS 2000a. Similarly, employment increased abruptly between OHS 1998 and LFS 2000a. Throughout these three sections, it was mentioned that these abrupt changes were either real (due to what happened to the economy during the time), due to the improvement of questionnaire design to capture the labour market aggregates better, as a result of the changes in labour market status derivation methodology throughout the years, or due to the adoption of different weighting techniques across the surveys. Unless the last three issues are addressed thoroughly, using all available labour surveys to analyse the labour market aggregates does not mean that the best labour market estimates and trends are derived. In other words, the incomparability and inconsistency issues mentioned above must be addressed before more comparable and reliable labour market trends could be derived.

It was mentioned in Section 4.4 that informal sector employment and self-employment fluctuations until the early LFSs might have caused the abrupt changes in the employment aggregates. Hence, Section 4.6 begins by investigating whether the employment trends become more reliable once these people are excluded, before the OHS/LFS/QLFS non-agricultural employment is compared with other employment data released by Stats SA. Next, labour market trends are re-examined after re-weighting the all surveys by the minimum cross entropy (CE) approach. Finally, the same labour market status derivation methodology (e.g., the LFS 2000b-LFS 2007 methodology) is applied to all surveys, if possible, to investigate if more comparable labour market estimates could be derived.

4.6.1 Comparison with other employment data

Figure 4.22 and Table 4.22 show the number of formal non-agricultural employees between 1997 and 2009[72]. After excluding the informal workers, self-employed as well as agricultural workers, the abrupt increase between OHS 1999 and LFS 2000a as seen in Figure 4.8 no longer happened. In fact, non-agricultural formal employment declined slightly between the two surveys (decreasing from 6.37 to 6.27 million). Furthermore, formal employment increased in a fair stable way between 1997 and 2008, except that there was a relatively rapid increase of 0.61 million between LFS 2007a and LFS 2007b (and this was attributed to the higher number of professionals in the latter survey – see the share of employed in the professionals occupational category suddenly increased from 4.7% in LFS 2007a to 7.6% in LFS 2007b in Table 4.15). Formal employment dropped by nearly 1 million between QLFS 2008Q4 and QLFS 2010Q2 due to the impact of recession, before increasing again. In fact, the level in QLFS 2009Q4 (8.51 million) was higher than what happened in QLFS 2008Q4 (8.87 million). To conclude, formal non-agricultural employment increased by 39.27% (from 6.11 million to 8.51 million) between 1997 and 2009.

Figure 4.22: Number of formal non-agricultural employees, 1997-2009



[72] It is not possible to derive informal sector employment in OHS 1994-OHS1996 – refer to footnote 69 and Figure 4.8.

Table 4.22: Number of formal non-agricultural employees (1 000s), 1997-2009

	Number (1 000s)	Absolute change (1 000s)	Percentage change
OHS 1997	6 107		
OHS 1998	6 152	46	0.7%
OHS 1999	6 374	222	3.6%
LFS 2000a	6 267	-107	-1.7%
LFS 2000b	6 599	332	5.3%
LFS 2001a	6 344	-255	-3.9%
LFS 2001b	6 562	217	3.4%
LFS 2002a	6 629	67	1.0%
LFS 2002b	6 684	55	0.8%
LFS 2003a	6 735	51	0.8%
LFS 2003b	6 854	119	1.8%
LFS 2004a	6 934	80	1.2%
LFS 2004b	7 168	234	3.4%
LFS 2005a	7 175	7	0.1%
LFS 2005b	7 462	287	4.0%
LFS 2006a	7 519	57	0.8%
LFS 2006b	7 816	297	4.0%
LFS 2007a	7 856	40	0.5%
LFS 2007b	8 471	615	7.8%
QLFS 2008Q1	8 694	223	2.6%
QLFS 2008Q2	8 781	89	1.0%
QLFS 2008Q3	8 767	-17	-0.2%
QLFS 2008Q4	8 874	107	1.2%
QLFS 2009Q1	8 763	-111	-1.2%
QLFS 2009Q2	8 712	-51	-0.6%
QLFS 2009Q3	8 455	-257	-3.0%
QLFS 2009Q4	8 505	50	0.6%

With regard to the other data sources on the number of non-agricultural employees, since 1998, Stats SA has been releasing employment statistics by conducting the enterprise-based surveys, namely the Survey of Employment and Earnings (SEE) and Quarterly Employment Statistics (QES). SEE was introduced in 1998. It took place on a quarterly basis, and the employment figures were derived from a sample of just above 10 000 private and public enterprises in the formal non-agricultural business sector. It covered VAT registered enterprises with an annual turnover of R300 000 or more in nominal terms (Stats SA 2001c). The only exception is that employment information in the mining and quarrying industry came from the Minerals Bureau of the Department of Minerals and Energy.

However, a drawback of the survey is that it did not collect information from various industries[73]. As a result, Stats SA, in collaboration with the South African Revenue Services

[73] These industries include the following (Altman 2008:128): agriculture, hunting, forestry and fishing; restaurants and other eating and drinking places, boarding houses, caravan parks and guest farms; storage, water and air transport; telecommunication services; financial institutions other than banking institutions and insurance companies; real estate and business services; educational services; medical, dental and other health services; welfare organizations; religious organizations; and recreational and cultural services.

(SARS) and the Department of Trade and Industry and Labour, re-engineered the register of businesses in 2002, which eventually formed the basis of the new sampling frame. This new 2002 sampling frame covered all the industries in the formal non-agricultural business sectors, that is, all previously excluded non-agricultural industries were covered by the 2002 sample (Stats SA 2003e). Stats SA released employment statistics with SEE for the last time in the first quarter of 2005.

Since the last quarter of 2004, Stats SA introduced the QES, which eventually replaced SEE from the second quarter of 2005. QES takes place on a quarterly basis, covering a sample of more than 20 000 private and public enterprises in the formal non-agricultural business sector (Stats SA 2006e). The information received is used to estimate employment and gross earnings that are used as inputs to the Gross Domestic Product (GDP) and to estimate key economic statistics on mean monthly earnings. A new sampling frame was introduced in 2006 (Stats SA 2006e). As in the SEE, the employment information in the mining and quarrying industry in the QES came from the Minerals Bureau of the Department of Minerals and Energy.

Table 4.23 and Figure 4.23 present the formal employment figures in each SEE and QES. The change in the SEE sampling frame led to a break in the series, with a very rapid increase in employment in the financial intermediation, insurance, real estate and business services (this is due to the fact that the old 1998 sampling frame excluded the capturing of those employed in real estate and business services), and a relatively minor abrupt increase of employment in the wholesale and retail as well as community, social and personal services industries[74].

Similarly, there was a sudden increase in the QES formal employment in the second quarter of 2006 due to the introduction of the new sampling frame, which led to an abrupt but moderate increase in employment in the manufacturing, wholesale and retail, financial intermediation, insurance, real estate and business services, as well as the community, social and personal services industries[75].

[74] How the adoption of the new sampling frame in 2002 affected the SEE employment in each industry falls beyond the scope of this dissertation and will not be discussed in further detail.
[75] How the adoption of the new sampling frame in 2006 affected the QES employment in each industry falls beyond the scope of this dissertation and will not be discussed in further detail.

Table 4.23: Number of formal non-agricultural employees in SEEs and QESs (1 000s), 1998Q2-2009Q4

	SEE (1998 sampling frame)	SEE (2002 sampling frame)	QES (2004 sampling frame)	QES (2006 sampling frame)
1998Q2	4 978			
1998Q3	4 963			
1998Q4	4 919			
1999Q1	4 928			
1999Q2	4 886			
1999Q3	4 842			
1999Q4	4 811			
2000Q1	4 777			
2000Q2	4 740			
2000Q3	4 707			
2000Q4	4 714			
2001Q1	4 674			
2001Q2	4 660			
2001Q3	4 649			
2001Q4	4 649			
2002Q1	4 632			
2002Q2	4 646			
2002Q3	4 679	6 510		
2002Q4	4 704	6 517		
2003Q1		6 497		
2003Q2		6 336		
2003Q3		6 369		
2003Q4		6 425		
2004Q1		6 447		
2004Q2		6 492		
2004Q3		6 600		
2004Q4		6 559	7 097	
2005Q1		6 560	6 945	
2005Q2			7 078	
2005Q3			7 165	
2005Q4			7 248	
2006Q1			7 238	
2006Q2			7 285	8 059
2006Q3			7 338	8 124
2006Q4				8 222
2007Q1				8 243
2007Q2				8 289
2007Q3				8 343
2007Q4				8 411
2008Q1				8 417
2008Q2				8 457
2008Q3				8 490
2008Q4				8 512
2009Q1				8 326
2009Q2				8 241
2009Q3				8 144
2009Q4				8 163

Figure 4.23: Number of formal non-agricultural employees in SEEs and QESs, 1998Q2-2009Q4

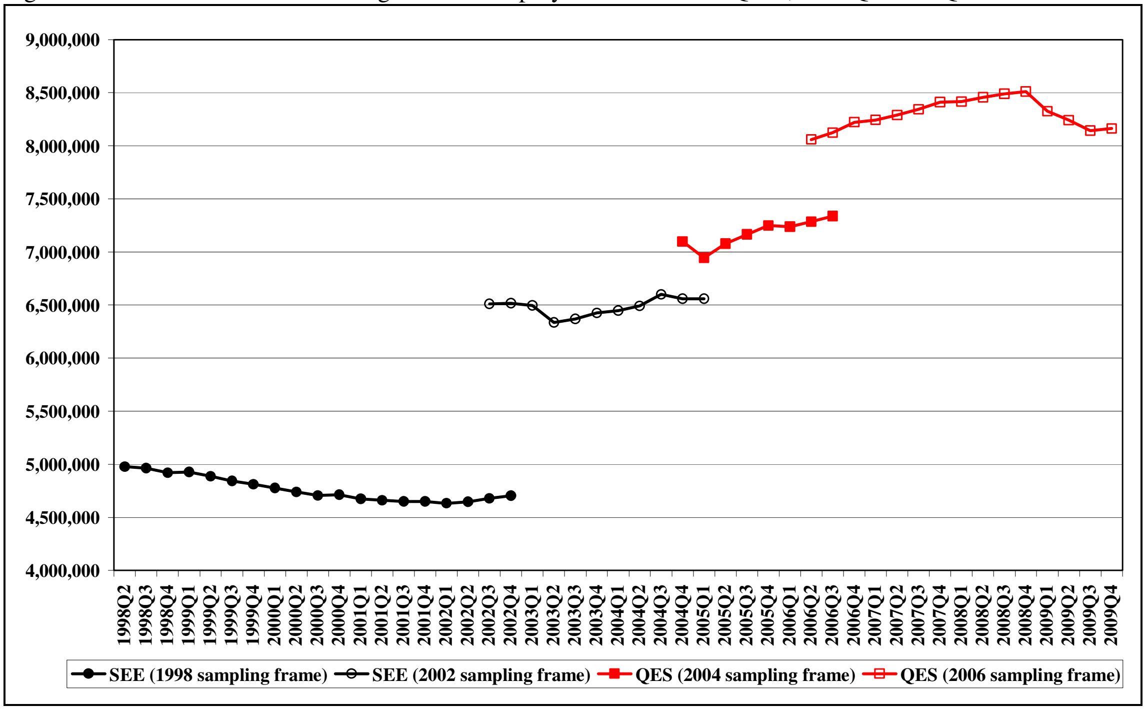


Figure 4.24: Number of formal non-agricultural employees in OHSs, LFSs, QLFSs, SEEs and QESs, 1997Q4-2009Q4

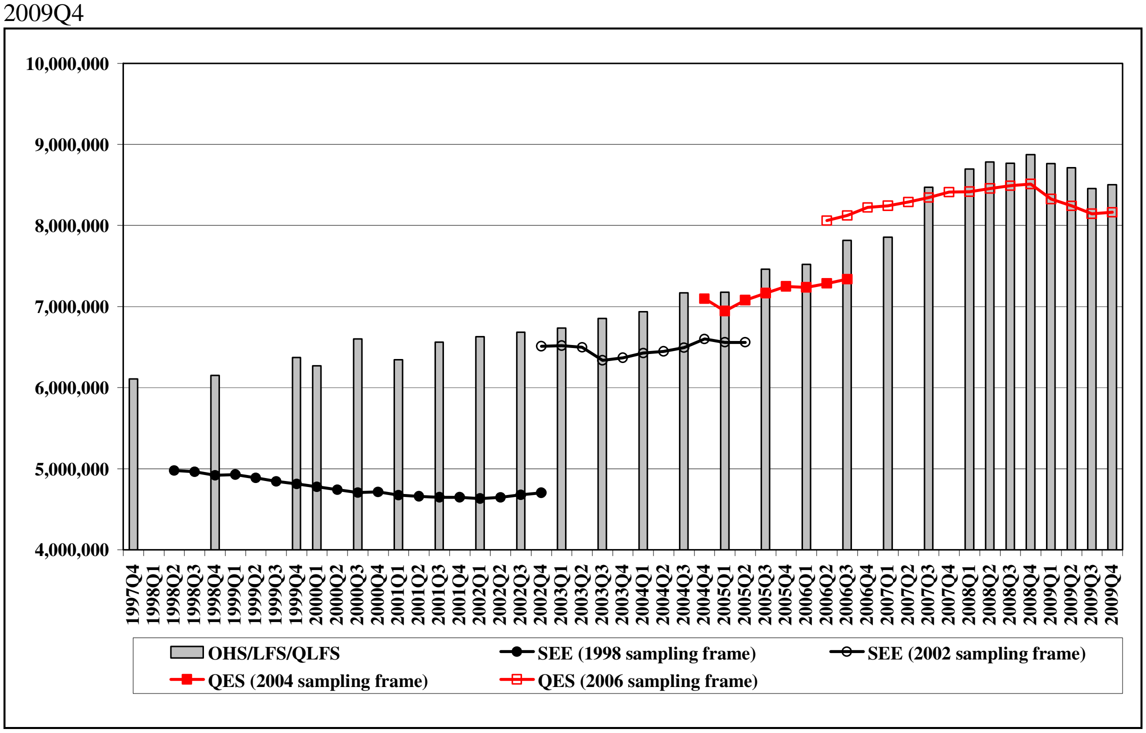


Figure 4.24 compares the OHS/LFS/QLFS non-agricultural employment with those derived by SEEs and QESs. It could be seen that the SEE (using the 2002 sampling frame) and QES (for both 2004 and 2006 sampling frames) aggregates and the LFS/QLFS aggregates have extremely similar trends in 2003-2009, except that the LFS/QLFS aggregates were always

greater between 2002 and 2009. Furthermore, it is interesting that between 1998 and 2001, the SEE (suing the 1998 sampling frame) showed that formal non-agricultural employment decreased (from 4.98 million to 4.65 million), but the opposite trend took place when looking at OHS 1998-LFS 2001 (formal non-agricultural employment increased from 6.15 million to 6.56 million). However, as mentioned in footnote 73, employment in quite a number of industries was not captured by the SEE 1998 sampling frame, so the SEE and OHS/LFS data might not be fully comparable between 1998 and 2001.

To conclude, the reliability and comparability of the OHS/LFS/QLFS data seems to have improved if only non-agricultural employees are examined. Furthermore, the LFS/QLFS non-agricultural employment data and the SEE/QES data since 2003 showed very similar trends. The results suggest that the LFS/QLFS non-agricultural employment aggregates are valid and reliable. In other words, the volatile employment data on informal sector employment, self-employment and agricultural employment might have caused the fluctuation of overall employment estimates in the OHSs and between OHS 1999 and LFS 2000a as discussed in Section 4.4.

4.6.2 Cross entropy re-weighting approach

This section re-examine the labour market trends after re-weighting the OHS 1994-QLFS 2009Q4 data by the minimum cross entropy (CE) approach, as discussed in Section 3.10. First, Table 4.24 presents information on the total population and working-age population using the original Stats SA weights and CE weights respectively. First, after the CE weights were used, the total population was higher in OHS 1995 – LFS 2007b (the population was about 1.0-1.5 million higher in OHS 1995 – OHS 1999, and 0.1-0.6 million higher in the LFSs) but slightly lower (by between 0.10-0.15 million) in the QLFSs. In addition, with the exception of OHS 1994, the working-age population was always higher in all surveys after using the CE weights.

Table 4.24: Total population and working-age population (1 000s), before and after the cross entropy approach was applied

	Total population		Working-age population	
	Stats SA weight	CE weight	Stats SA weight	CE weight
OHS 1994	40 251	40 262	24 075	24 032
OHS 1995	39 660	41 207	24 191	24 877
OHS 1996	40 583	42 154	24 909	25 659
OHS 1997	41 443	42 890	25 506	26 230
OHS 1998	42 212	43 650	25 665	26 885
OHS 1999	43 272	44 363	26 247	27 431
LFS 2000a	43 620	44 587	26 465	27 681
LFS 2000b	44 821	44 976	27 836	27 933
LFS 2001a	45 080	45 291	28 062	28 286
LFS 2001b	45 081	45 600	28 084	28 580
LFS 2002a	45 325	45 890	28 298	28 852
LFS 2002b	45 561	46 172	28 495	29 101
LFS 2003a	45 810	46 433	28 725	29 351
LFS 2003b	46 046	46 685	28 906	29 556
LFS 2004a	46 271	46 915	29 100	29 767
LFS 2004b	46 490	47 139	29 271	29 946
LFS 2005a	46 700	47 346	29 490	30 158
LFS 2005b	46 917	47 549	29 663	30 295
LFS 2006a	47 184	47 739	29 818	30 451
LFS 2006b	47 429	47 925	29 973	30 561
LFS 2007a	47 652	48 100	30 161	30 717
LFS 2007b	47 883	48 272	30 387	30 884
QLFS 2008Q1	48 489	48 407	30 764	30 995
QLFS 2008Q2	48 589	48 490	30 875	31 078
QLFS 2008Q3	48 687	48 571	30 950	31 130
QLFS 2008Q4	48 780	48 647	31 047	31 190
QLFS 2009Q1	48 873	48 724	31 145	31 260
QLFS 2009Q2	48 968	48 803	31 245	31 333
QLFS 2009Q3	49 060	48 879	31 325	31 393
QLFS 2009Q4	49 148	48 951	31 411	31 451

Note: Since the QLFS took place during a 3-month period, the February population figure derived by the ASSA model was used to derive the CE weights in the Q1 survey. Similarly, the May, August and November ASSA model's population figures were used to derive the CE weights in the Q2, Q3 and Q4 surveys respectively.

Table 4.25 and Figures 4.25-4.31 show the number of LF, employed and unemployed, as well as the LFPRs and unemployment rates when using the Stats SA weights and CE weights respectively. First, the narrow LF was always slightly higher when using the CE weights. In addition, the trends on the LF remained the same after using the CE weights, i.e., the abrupt increase between OHS 1996 and OHS 1999, followed by a very rapid increase during the changeover between OHS and LFS; a slight upward trend took place in the LFSs, before the LF declined in QLFS 2009 due to the impact of the recession. The only difference was that LF decreased from OHS 1994 and OHS 1995 when using the Stats SA weights[76], but the opposite took place when using the CE weights (Figure 4.25).

[76] Note that OHS 1994 and OHS 1995 were weighted with Census 1991 and Census 1996 weights respectively, as mentioned in Section 2.4.1.

Table 4.25: Selected statistics on labour market trends, before and after the cross entropy approach was applied

	Narrow labour force (1 000s)		Employed (1 000s)		Narrow unemployed (1 000s)	
	Stats SA weight	**CE weight**	**Stats SA weight**	**CE weight**	**Stats SA weight**	**CE weight**
OHS 1994	11 884	11 925	9 439	9 511	2 445	2 415
OHS 1995	11 528	12 097	9 499	10 151	2 028	1 946
OHS 1996	11 191	11 792	8 966	9 533	2 224	2 259
OHS 1997	11 545	12 091	9 094	9 603	2 451	2 488
OHS 1998	12 528	13 407	9 370	10 151	3 158	3 257
OHS 1999	13 510	14 306	10 356	11 068	3 154	3 238
LFS 2000a	16 228	17 164	11 874	12 723	4 331	4 442
LFS 2000b	16 444	16 676	12 224	12 481	4 157	4 196
LFS 2001a	16 668	16 963	12 260	12 526	4 408	4 437
LFS 2001b	15 818	16 308	11 168	11 566	4 650	4 742
LFS 2002a	16 494	16 991	11 603	11 995	4 891	4 996
LFS 2002b	16 215	16 744	11 284	11 718	4 931	5 026
LFS 2003a	16 409	16 985	11 298	11 759	5 111	5 227
LFS 2003b	15 841	16 430	11 411	11 881	4 429	4 549
LFS 2004a	15 788	16 390	11 378	11 854	4 410	4 537
LFS 2004b	15 761	16 324	11 630	12 103	4 131	4 221
LFS 2005a	16 173	16 747	11 894	12 408	4 278	4 339
LFS 2005b	16 770	17 388	12 288	12 823	4 482	4 564
LFS 2006a	16 708	17 296	12 438	12 926	4 270	4 371
LFS 2006b	17 173	17 698	12 787	13 248	4 386	4 450
LFS 2007a	16 966	17 579	12 635	13 176	4 331	4 403
LFS 2007b	17 194	17 711	13 293	13 756	3 901	3 955
QLFS 2008Q1	17 826	18 162	13 637	13 961	4 189	4 201
QLFS 2008Q2	17 864	18 172	13 749	14 059	4 115	4 113
QLFS 2008Q3	17 789	18 093	13 669	13 958	4 120	4 135
QLFS 2008Q4	17 733	18 018	13 862	14 135	3 871	3 883
QLFS 2009Q1	17 833	18 120	13 653	13 936	4 181	4 184
QLFS 2009Q2	17 511	17 777	13 388	13 667	4 123	4 110
QLFS 2009Q3	17 086	17 328	12 897	13 163	4 190	4 165
QLFS 2009Q4	17 146	17 364	12 984	13 231	4 162	4 133

Figure 4.25: Narrow labour force before and after the cross entropy approach was applied, 1994-2009

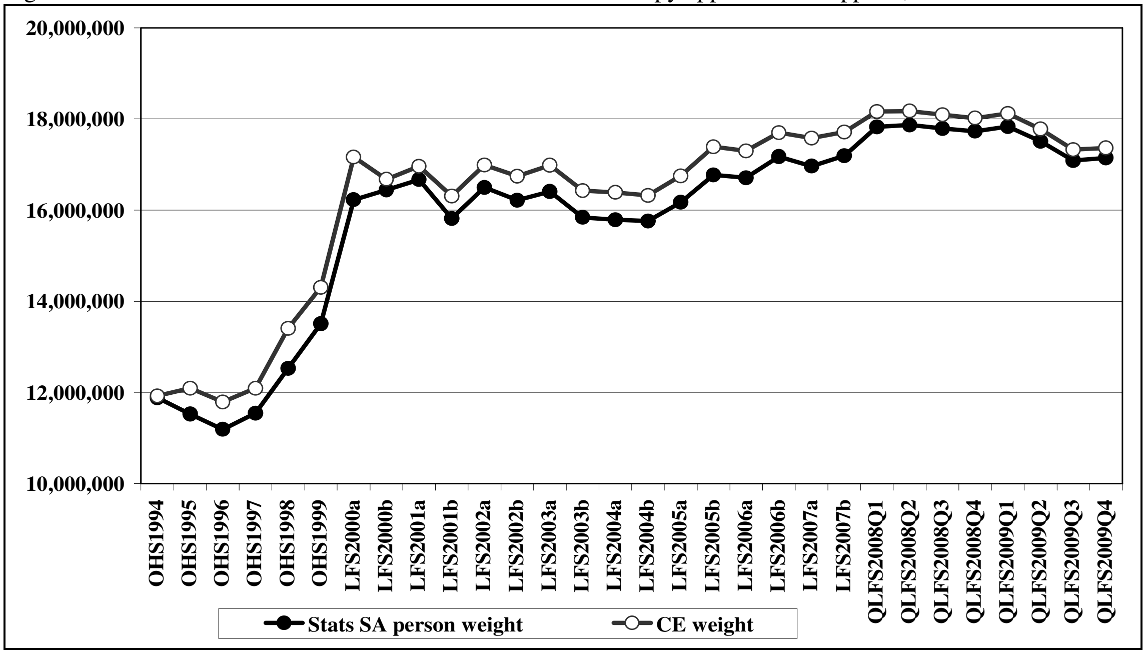
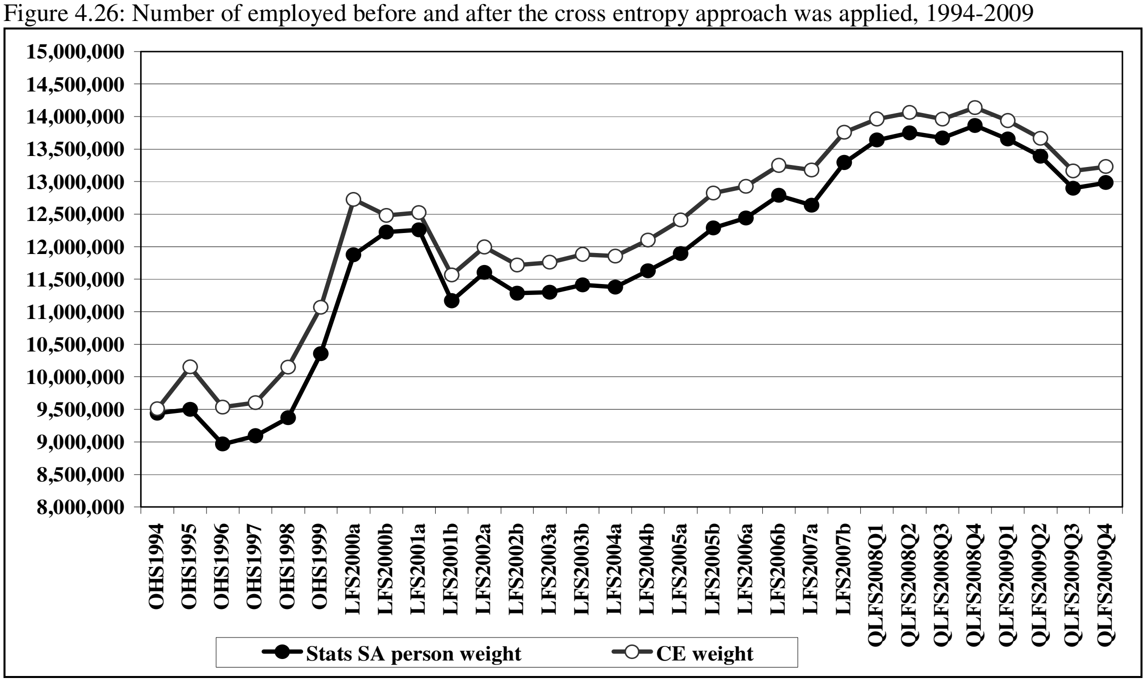


Figure 4.26 shows that the number of employed was greater in all surveys after using the CE weights. However, the employment trends remained unchanged after the application of the CE weights, i.e., the OHS 1995 employment was still higher when compared with the OHS 1994 and 1996-1998 numbers, before an abrupt increase took place between OHS 1998 and OHS 1999, followed by an even greater increase in LFS 2000a. Also, the sudden one million decrease in employment between LFS 2001a and LFS 2001b, as well as the upward trend between 2002 and 2008 remained unchanged, after applying the CE weights.

Figure 4.26: Number of employed before and after the cross entropy approach was applied, 1994-2009



Figures 4.27 and 4.28 provide more information by showing employment by gender as well as black employment respectively, and the results once again showed that the application of the CE weights did not result in more consistent employment trends.

Figure 4.27: Number of employed by gender before and after the cross entropy approach was applied, 1994-2009

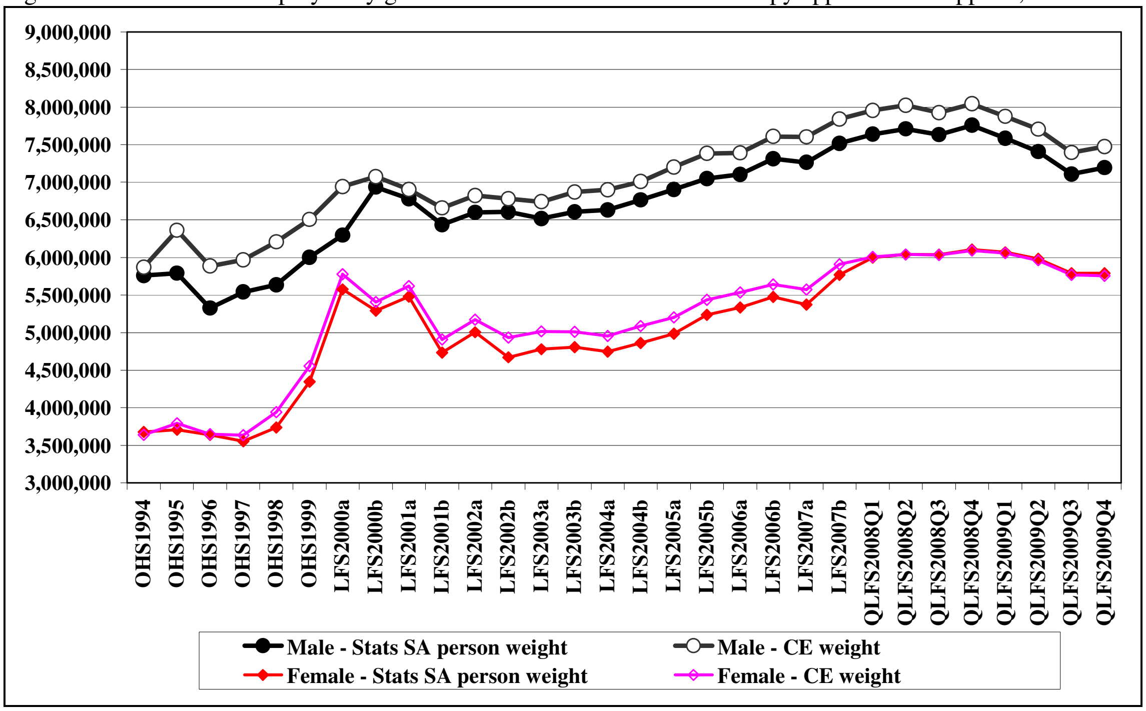


Figure 4.28: Number of black employed before and after the cross entropy approach was applied, 1994-2009

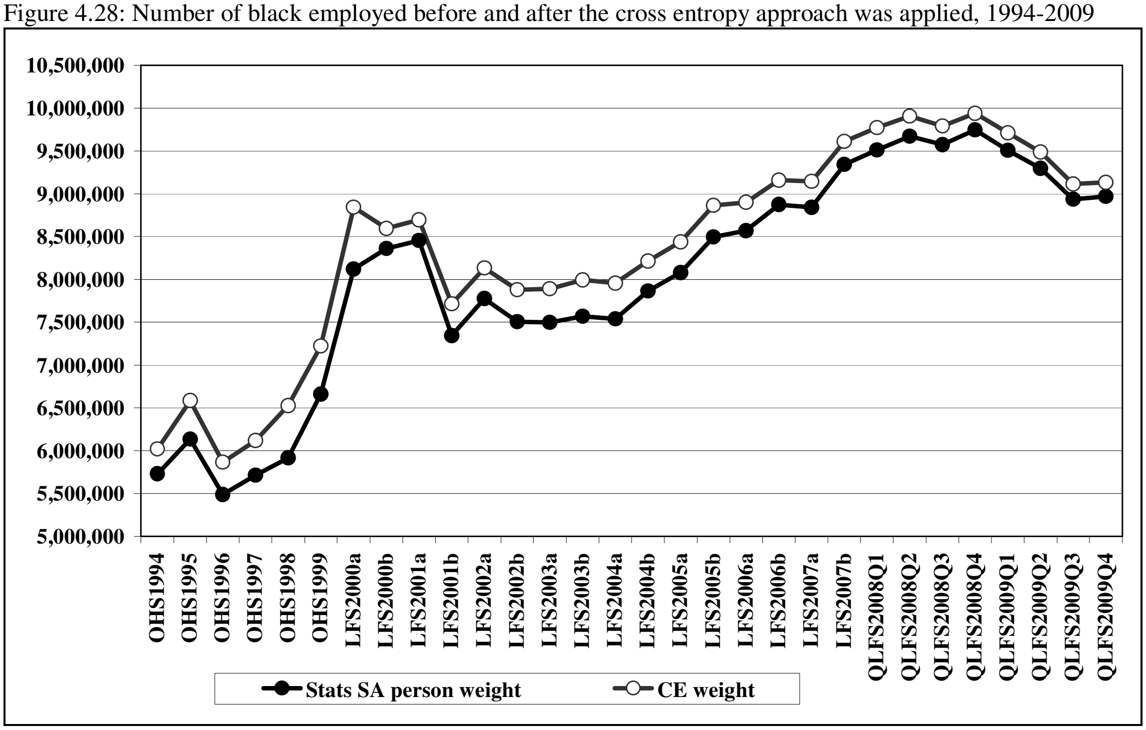
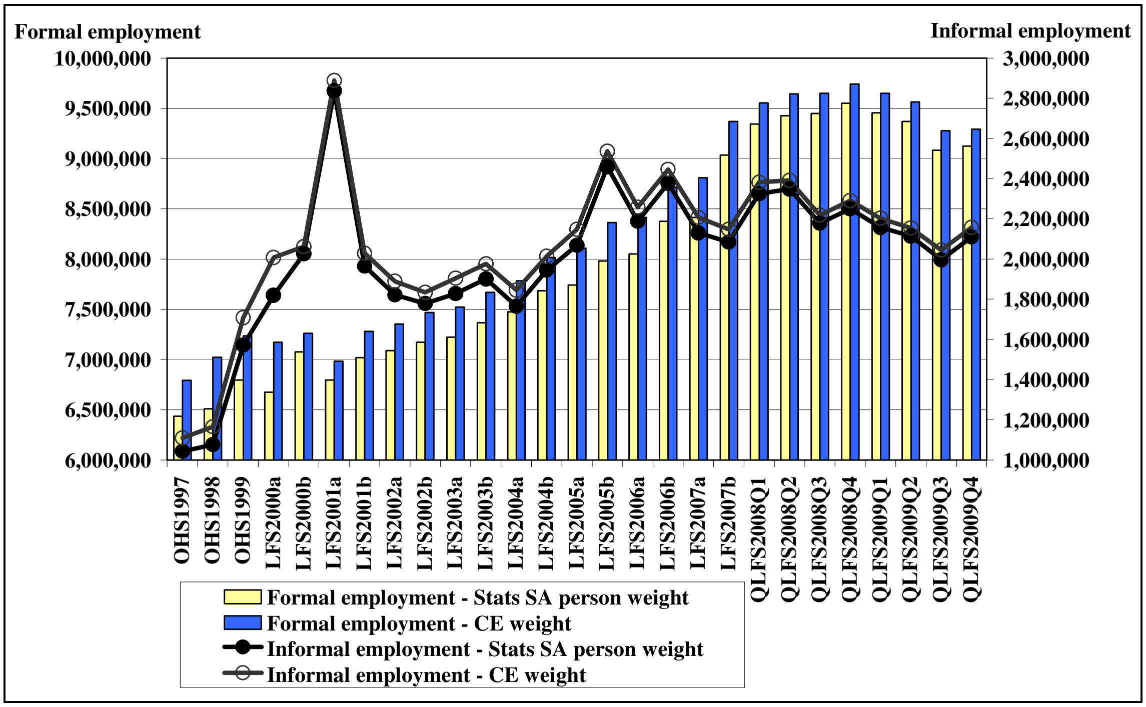


Figure 4.29 shows the formal and informal employment using the Stats SA and CE weights, and it can be seen that the use of the CE weights did not reduce the large number of informal employed found in LFS 2001a.

Figure 4.29: Formal and informal employment before and after the cross entropy approach was conducted, 1997-2009



The narrow unemployed numbers using Stats SA weights and CE weights were very close (See Figure 4.30), and the use of the CE weights did not result in a smoother unemployment trend (e.g., the nearly 1.2 million increase of unemployed between OHS 1999 and LFS 2000a was still observed). Finally, although the narrow unemployment rates were slightly lower when using the CE weights (Figure 4.31), the trends using the two weights were very similar.

Figure 4.30: Number of narrow unemployed before and after the cross entropy approach was applied, 1994-2009

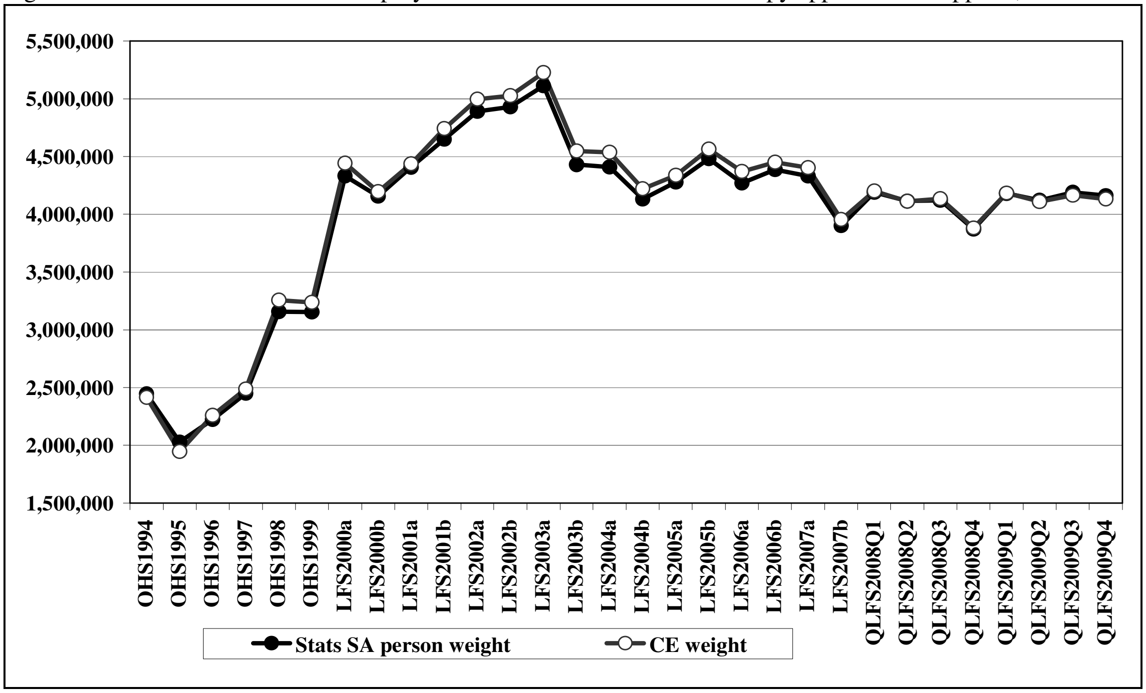
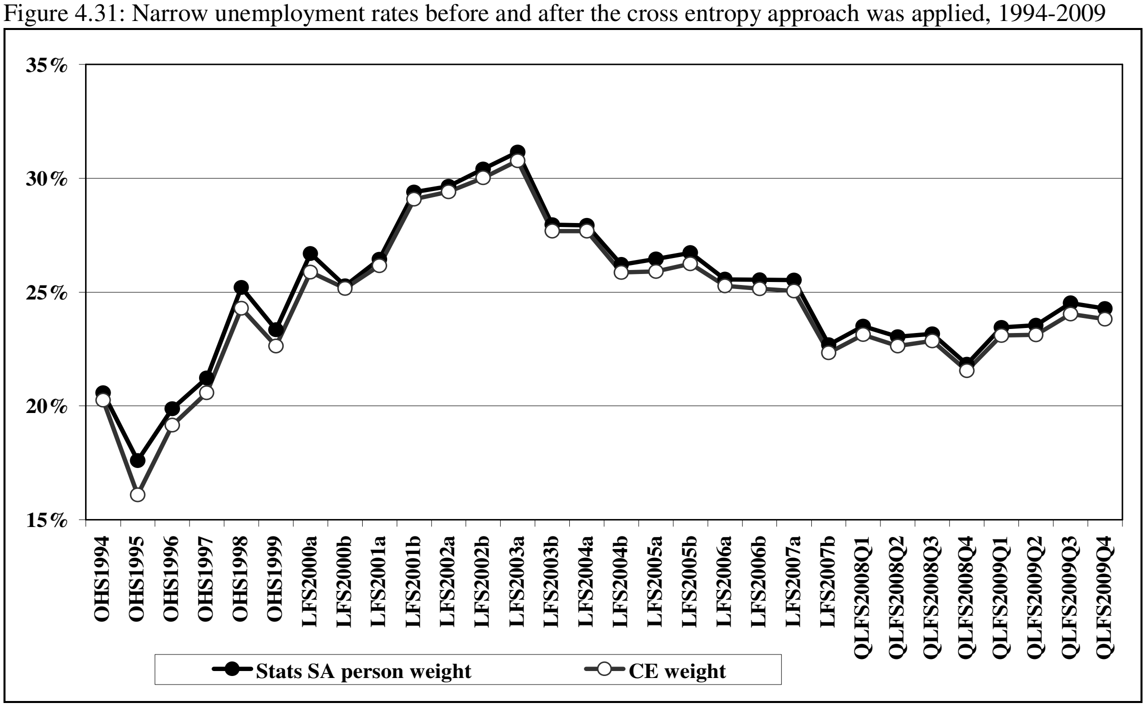


Figure 4.31: Narrow unemployment rates before and after the cross entropy approach was applied, 1994-2009



To conclude, the use of the CE weights clearly did not result in any changes in the trends in LF, employed, unemployed and unemployment rates. This implies that the abrupt changes observed between surveys (e.g., the sudden increase between OHS 1999 and LFS 2000a) were unlikely to be caused by post-stratification weighting errors, but rather caused by the fact that either these shifts were real, or the LF / employed / unemployed were better captured in the LFSs and QLFSs as a result of the improvement of the questionnaire design, or rather to the changes in the methodology to derive labour market status. The latter issue is the focus of Section 4.6.3.

4.6.3 Application of a consistent labour market status derivation methodology across surveys

Section 3.10 discussed that lots of changes took place between OHS 1996-LFS 2000a with regard to the labour market status derivation methodology, before a consistent approach was adopted in LFS 2000b-LFS 2007b. It is possible that the abrupt changes in the labour market aggregates in the LFSs as well as the very sudden increase of these aggregates between OHS 1999 and LFS 2000a took place mainly due to the changes in the methodology to derive labour market status. Furthermore, a new methodology was used since the launch of QLFS in 2008, but this methodology is very different from the LFS 2000b-LFS 2007b methodology, resulting in much smaller number of unemployed in broad terms. Hence, this section aims to examine the two questions:

- What would have happened to the labour market aggregates in the OHSs and LFS 2000a, had the LFS 2000b-LFS 2007b methodology been applied?
- What would have happened to labour market aggregates in the OHSs and LFSs, had the latest QLFS methodology been applied?

As the questions and the methodology to derive the employed have changed drastically across the surveys (see Table 3.6), it is not possible to apply a consistent methodology (e.g., QLFS methodology) in all surveys to find out the number of employed had the same methodology been used throughout the years. The focus of Section 4.6.3 would rather be on unemployment.

4.6.3.1 Application of the LFS 2000b-LFS 2007b methodology in the earlier surveys

Section 3.11 discussed in detail how the LFS 2000b-LFS 2007b methodology differs from the methodologies used in the OHSs as well as LFS 2000a. The former methodology, however, could not be applied in OHS 1994-1995 due to the incomparability of the questions asked to derive the unemployed.

Table 4.26 presents what would have happened to the number of unemployed and unemployment rate in OHS 1996-LFS 2000a, had the LFS 2000b-LFS 2007b methodology been applied in these surveys[77]. The results show that these aggregates had increased in all these surveys, except the slight decrease in LFS 2000a in broad terms. The increase was relatively greater in OHS 1996, OHS 1998 and OHS 1999. In particular, the number of broad unemployed and broad unemployment rate in OHS 1999 increased most rapidly (by 658 000 and 4.1 percentage points respectively) after the adoption of the LFS 2000b-LFS 2007 methodology.

Furthermore, as discussed in Section 4.5 (See Table 4.19 and Figure 4.13), using the original methodology, there was a rapid increase of the number of narrow unemployed by 1.18 million (from 3.15 million to 4.33 million) and broad unemployed by 0.68 million (from 5.88 million to 6.55 million) during the changeover between the OHS and LFS. Had the consistent LFS 2000b-LFS 2007b methodology been applied to these surveys, the abovementioned mentioned abrupt increase became slightly less serious in the narrow terms (increasing by only 0.91 million instead of 1.18 million previously) but no longer happened in broad terms. In fact, looking at the latter result in greater detail, the number of broad unemployed showed a slight decline of 7 700 (decreasing from 6.53 million to 6.46 million).

Table 4.26: Unemployment in OHS 1996-LFS 2000a using different labour market status methodologies

	[A]: Original methodology in each survey		[B]: LFS 2000b-LFS 2007b methodology		Difference: [B] – [A]	
	Narrow	Broad	Narrow	Broad	Narrow	Broad
	Number of unemployed (1 000s)					
OHS 1996	2 224	4 566	2 525	5 069	301	503
OHS 1997	2 451	5 202	2 560	5 256	109	54
OHS 1998	3 158	5 626	3 406	6 122	248	496
OHS 1999	3 154	5 875	3 451	6 533	297	658
LFS 2000a	4 331	6 550	4 365	6 456	33	-94
	Unemployment rate					
OHS 1996	19.9%	33.7%	22.6%	37.5%	2.7%	3.7%
OHS 1997	21.2%	36.4%	22.2%	36.8%	0.9%	0.4%
OHS 1998	25.2%	37.5%	27.2%	40.8%	2.0%	3.3%
OHS 1999	23.3%	36.2%	25.5%	40.2%	2.2%	4.1%
LFS 2000a	26.7%	35.5%	26.9%	35.0%	0.2%	-0.5%

To conclude, it seems the changes in the labour market status derivation methodology in the OHSs and LFS 2000a could partly explain the fluctuations and abrupt changes in the unemployment aggregates in these surveys. Also, the unemployment aggregates could have been under-estimated, with the exception of LFS 2000a unemployment in broad terms. However, even after adopting the LFS 2000b-LFS 2007 methodology in the OHSs and LFS 2000a, the abrupt increase of narrow unemployed between OHS 1999 and LFS 2000a still took place. This implies this sudden increase was either real (due to what happened to the

[77] Table B.1 in Appendix B shows the Stata do-file to derive the unemployed in OHS 1996-LFS 2000a.

economy) or due to the big changes in the questionnaire design. The analyses in Section 4.6.2 already found that this sudden increase was not caused by different weighting techniques.

4.6.3.2 Application of the QLFS methodology in the earlier surveys

The discussion in Section 3.10 as well as the findings in Section 4.5 showed that the QLFS methodology in broad terms is not comparable with the broad methodologies in the OHSs and LFSs. Hence, Section 4.6.3.2 deals with the issue of what would have happened to the unemployment aggregates, in particular in broad terms, had the QLFS methodology been applied on the OHSs and LFSs. Unfortunately this is not possible in the OHSs due to the drastic changes in the categorisation of the answers of the questions used to derive labour market status. In addition, the question '3.8: What was the main reason why you did not try to find work or start a business in the last four weeks?' which was used in the broad methodology in the QLFS, was not asked in the OHSs.

Despite the fact that the QLFS methodology could be applied on the LFSs, it requires minor adjustment, because the QLFS methodology concerns about whether the labour force is ready to accept a job offer or to start a business within one week, but the LFS methodology was only concerned about the acceptance of a job offer within one week. In fact, the questions on whether the respondents were ready to start a business were not asked in the LFSs. Hence, the QLFS methodology is revised slightly in the way that it does not take the respondents' answers on the readiness to start a business into consideration when deriving their labour market status, before it could be applied on the LFSs. For the remainder of Section 4.6.3.2, this will be referred to as the 'revised QLFS methodology'.

Table 4.27, Figures 4.32 and Figure 4.33 present the unemployment aggregates in the LFSs and QLFSs after using the revised QLFS methodology[78]. The number of unemployed and unemployment rates in both narrow and broad terms in the QLFSs only showed negligible changes, after the revised methodology is used (e.g., the two line charts almost overlap each other in Figures 4.32 and 4.33). This suggests that by ignoring the respondents' answers on the readiness to start a business in the revised methodology, the QLFS labour aggregates are not affected significantly. In contrast, the use of the revised QLFS methodology caused slight changes in the narrow unemployment aggregates in the LFSs, with the exception that these aggregates showed relatively greater increase in the 2000-2001 LFSs (e.g., the number of narrow unemployed increased by between 46 000 and 90 000, and the narrow unemployment

[78] Table B.2 in Appendix B shows the Stata do-file to derive the unemployed in OHS 1996-LFS 2000a.

rate increased by between 0.3 and 0.6 percentage points, as shown in Table 4.27).

Table 4.27: Unemployment in the LFSs and QLFSs using the revised QLFS methodology

	[A]: Original methodology in each survey		[B]: Revised QLFS methodology		Difference: [B] – [A]	
	Narrow	Broad	Narrow	Broad	Narrow	Broad
	Number of unemployed (1 000s)					
LFS2000a	4 331	6 550	4 395	5 747	64	-803
LFS2000b	4 157	6 372	4 244	5 738	87	-633
LFS2001a	4 408	7 101	4 454	6 255	46	-847
LFS2001b	4 650	7 640	4 740	6 797	90	-844
LFS2002a	4 891	7 932	4 887	6 915	-4	-1 018
LFS2002b	4 931	8 121	4 969	7 023	38	-1 098
LFS2003a	5 111	8 345	5 125	7 239	14	-1 105
LFS2003b	4 429	8 198	4 430	6 842	1	-1 356
LFS2004a	4 410	8 172	4 394	6 754	-16	-1 418
LFS2004b	4 131	8 074	4 113	6 697	-18	-1 377
LFS2005a	4 278	8 098	4 297	6 659	19	-1 439
LFS2005b	4 482	7 791	4 498	6 851	16	-940
LFS2006a	4 270	7 949	4 318	6 873	48	-1 076
LFS2006b	4 386	7 599	4 350	6 555	-37	-1 044
LFS2007a	4 331	7 830	4 348	6 901	18	-929
LFS2007b	3 901	7 340	3 936	6 369	35	-970
QLFS2008Q1	4 189	5 366	4 160	5 335	-29	-31
QLFS2008Q2	4 115	5 193	4 093	5 169	-22	-23
QLFS2008Q3	4 120	5 191	4 103	5 173	-17	-18
QLFS2008Q4	3 871	5 039	3 850	5 019	-20	-20
QLFS2009Q1	4 181	5 396	4 163	5 378	-17	-17
QLFS2009Q2	4 123	5 639	4 097	5 612	-26	-27
QLFS2009Q3	4 190	5 820	4 168	5 796	-22	-24
QLFS2009Q4	4 162	5 847	4 145	5 830	-17	-17
	Unemployment rate					
LFS2000a	26.7%	35.5%	27.1%	31.2%	0.4%	-4.4%
LFS2000b	25.3%	34.1%	25.8%	30.8%	0.5%	-3.4%
LFS2001a	26.4%	36.7%	26.7%	32.3%	0.3%	-4.4%
LFS2001b	29.4%	40.6%	30.0%	36.1%	0.6%	-4.5%
LFS2002a	29.7%	40.6%	29.6%	35.4%	0.0%	-5.2%
LFS2002b	30.4%	41.8%	30.6%	36.2%	0.2%	-5.7%
LFS2003a	31.1%	42.5%	31.2%	36.9%	0.1%	-5.6%
LFS2003b	28.0%	41.8%	28.0%	34.9%	0.0%	-6.9%
LFS2004a	27.9%	41.8%	27.8%	34.5%	-0.1%	-7.3%
LFS2004b	26.2%	41.0%	26.1%	34.0%	-0.1%	-7.0%
LFS2005a	26.5%	40.5%	26.6%	33.3%	0.1%	-7.2%
LFS2005b	26.7%	38.8%	26.8%	34.1%	0.1%	-4.7%
LFS2006a	25.6%	39.0%	25.8%	33.7%	0.3%	-5.3%
LFS2006b	25.5%	37.3%	25.3%	32.2%	-0.2%	-5.1%
LFS2007a	25.5%	38.3%	25.6%	33.7%	0.1%	-4.5%
LFS2007b	22.7%	35.6%	22.9%	30.9%	0.2%	-4.7%
QLFS2008Q1	23.5%	28.2%	23.3%	28.1%	-0.2%	-0.2%
QLFS2008Q2	23.0%	27.4%	22.9%	27.3%	-0.1%	-0.1%
QLFS2008Q3	23.2%	27.5%	23.1%	27.4%	-0.1%	-0.1%
QLFS2008Q4	21.8%	26.7%	21.7%	26.6%	-0.1%	-0.1%
QLFS2009Q1	23.4%	28.3%	23.3%	28.2%	-0.1%	-0.1%
QLFS2009Q2	23.5%	29.6%	23.4%	29.5%	-0.1%	-0.1%
QLFS2009Q3	24.5%	31.1%	24.4%	31.0%	-0.1%	-0.1%
QLFS2009Q4	24.3%	31.1%	24.2%	31.0%	-0.1%	-0.1%

Figure 4.32: Number of unemployed in the LFSs and QLFSs using the revised QLFS methodology

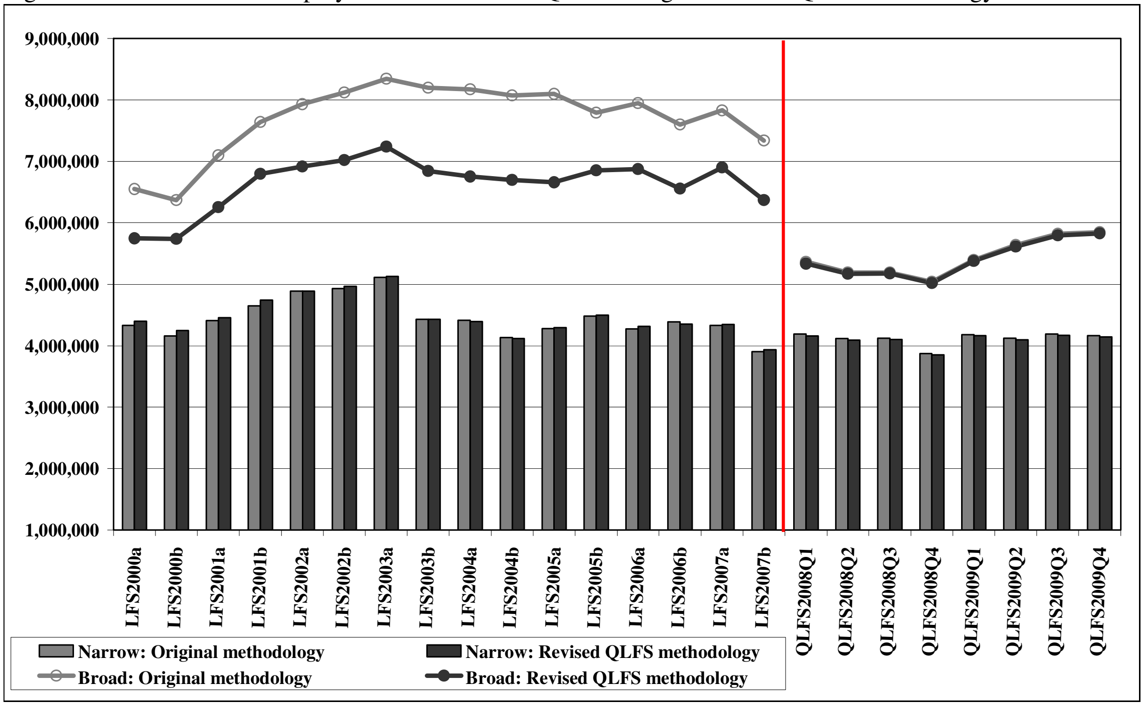
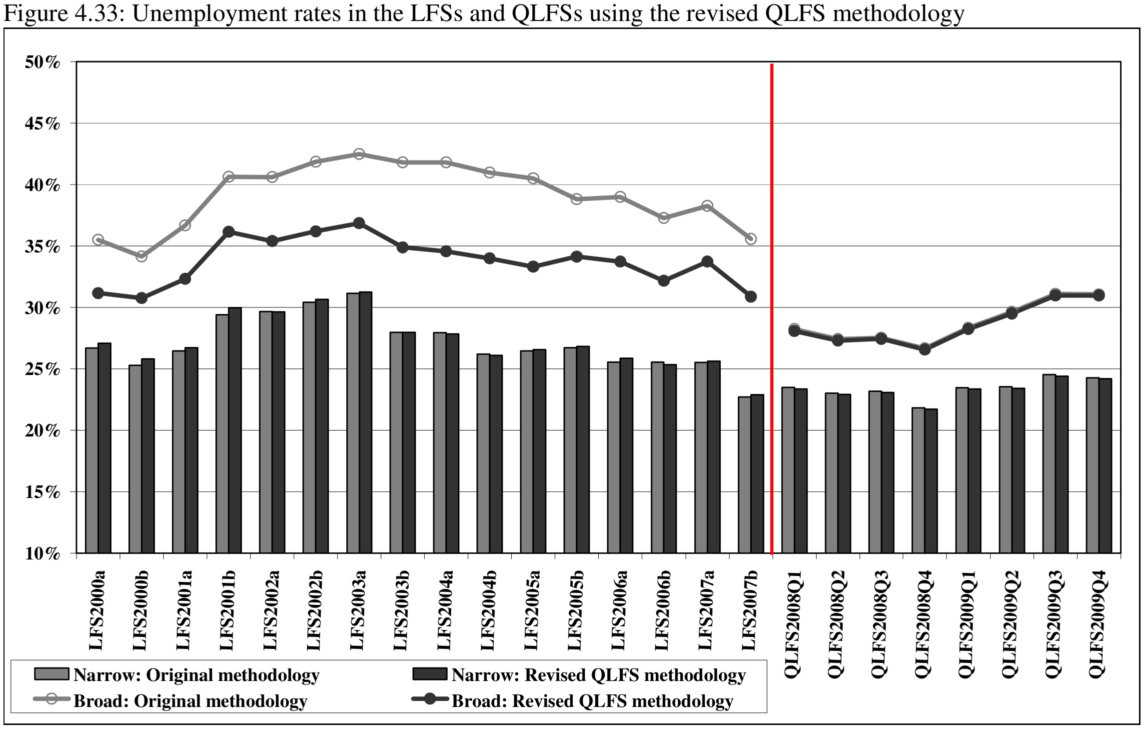


Figure 4.33: Unemployment rates in the LFSs and QLFSs using the revised QLFS methodology



However, after the adoption of the revised QLFS methodology, the number of broad unemployed decreased in the LFSs, by as little as 0.63 million in LFS 2000b and as much as 1.44 million in LFS 2005a. As a result of the lower number of broad unemployed in the LFSs after the use of this revised QLFS methodology, the extent of the abrupt decrease in the broad unemployed as well as broad unemployment rate became smaller during the changeover from

LFS to QLFS. This is indicated by the fact that the number of broad unemployed decreased drastically by 1.97 million (from 7.34 million to 5.37 million) between LFS 2007b and QLFS 2008Q1) if the original methodology was applied in each survey, but it only decreased by 1.03 million (from 6.37 million to 5.34 million) between the two surveys after the use of the revised QLFS methodology in both surveys. Similarly, the broad unemployment rate decreased abruptly by 7.4 percentage points (from 35.6% to 28.2%) between the two surveys previously, but the decline became smaller (by 2.8 percentage points – from 30.9% to 28.1%) after the use of the revised QLFS methodology.

Hence, the use of the revised QLFS methodology in all LFSs and QLFSs improve the comparability of the unemployment aggregates to a certain extent. However, the above findings still indicate an abrupt decrease (despite the extent of it being less serious) of these aggregates between LFS 2007b and QLFS 2008Q1. In other words, the incomparability issue across the two surveys was not fully solved as a result of the use of the revised QLFS methodology.

Thus, the possible reasons for the still relatively higher number of unemployed in broad terms in the LFSs could be real, or due to the difference in the questionnaire structure between LFSs and QLFSs. Looking at the latter factor in greater detail, it was found that the number of categories of the question on why the person did not work or start a business in the last four weeks were only 11 in LFSs but 16 in QLFSs. The proportions of respondents declaring the options 'No jobs available in the area' and 'lost hope of finding any kind of work'[79] were only about 48% and 5% respectively in the QLFSs, but were approximately 52% and 11% in the LFSs. So, this might have explained the higher number of broad unemployed in the LFSs even after the revised QLFS methodology was applied.

Furthermore, with regard to how soon the respondent could start working if being offered a job, the respondents were given the options 'within a week', 'within two weeks', 'within four weeks' and 'later than four weeks from now' to choose from, but this question was asked differently as whether the respondents could start working within a week if they are offered a job, and they could only have 'yes', 'no' and 'don't know' to choose from. Thus, the difference in the way this question was asked could also play a role to explain why the number of broad unemployed was still higher with the use of the revised QLFS methodology.

---

[79] These are two of the three options the respondents must declare before they are defined as discouraged workseekers (and eventually being included as part of broad unemployed) – refer to Figure 3.16.

4.7 Chapter summary

Given the importance of the labour market to the economic growth of any country, it is vital to infer trends from the available labour data correctly. In South Africa, several researchers compared selected household surveys with each other and then derived conclusions about the ‘trends’ in the labour market for the whole period between surveys, with the most commonly used approach being the comparison of OHS 1995 with the latest available OHS or LFS at the time of the study. Such a methodology may give misleading results and that it is preferable to look at all the available surveys before real trends could be determined. In fact, one must interpret the OHS 1994-1995 labour market results with caution, as the labour market status derivation methodology is not known (refer to Section 2.4.2.2).

Chapter 4 began by providing a literature review of the recent studies on the South African labour market trends, before investigating trends in LF, LFPRs and employment, work activities of the employed, as well as unemployment and unemployment from 1994 to 2009 by using all available OHSs, LFSs and QLFSs during the period. It was found that the LF and LFPR in both narrow and broad terms showed a rapid increase during the OHSs (with the exception of the slight decrease between 1994 and 1996), followed by a sudden increase during the changeover from OHS to LFS. The narrow LF and LFPR have since increased slightly, while the broad LF and LFPR have stabilised. The trends in the LFSs and QLFSs did not suggest that any feminisation of labour force had taken place after the OHS years, contrary to the findings of recent studies (which only compared OHS 1995 with an LFS).

The number of employed showed huge fluctuations, and it was only since LFS 2004b that the employment increased in a stable and continuous fashion. Hence, if different reference points are used in the calculation of TGR, AGR and EAR, one may draw contradictory conclusions regarding whether job creation or jobless growth occurred in the South African economy. Furthermore, both the narrow and broad unemployment rates increased from OHS 1994 to LFS 2003a, followed by a downward trend from LFS 2003b onwards. A rising trend was observed in 2009 due to economic recession.

Examining all available OHSs, LFSs and QLFSs, despite giving a more clearer picture on the possible fluctuations and trends of the labour market aggregates, does not suggest that the results are fully reliable and comparable. In fact, analysing all available labour surveys makes it easier to identify abrupt changes of the aggregates across some surveys, and these fluctuations need to be addressed, if possible, to improve the reliability and comparability of the labour market estimates. Hence, the last section of Chapter 4 first compared with the OHS/LFS/QLFS non-agricultural employment data with the SEE and QES data, and it was found that the LFS/QLFS employment aggregates became more stable and were very comparable with the SEE/QES data, once the unstable informal sector employment, agricultural employment and self-employment were excluded from the analyses.

In addition, after using the minimum cross entropy approach to re-weight all datasets under study, it was found that the trends in the number of narrow LF, employed and narrow unemployed as well the narrow unemployment rates did not show any changes. This implies that the large discrepancies in these labour market variables (especially between OHS 1994 and LFS 2000a) were not the result of using the inconsistent Stats SA weights. Finally, the LFS 2000b-LFS 20007b labour market status derivation methodology was applied in OHS 1996-LFS 2000a and it was found that the abrupt changes in the broad unemployment between OHS 1999 and LFS 2000a no longer happened. In contrast, when the revised QLFS methodology was applied in all LFSs and QLFSs, the number of broad unemployed became lower in the LFSs (despite still being relatively greater when compared with the QLFSs), and the extent of the sudden decrease of broad unemployed became less serious. Therefore, the results suggest that the comparability of the labour market aggregates across the surveys could be improved to a certain extent by applying a consistent labour market status derivation methodology in all surveys.

Having examined the labour market trends since the transition, the next two chapters focus on poverty and inequality trends using all available surveys.

CHAPTER FIVE: POVERTY AND INEQUALITY: CONCEPTS, MEASUREMENTS, DERIVATION OF INCOME AND EXPENDITURE VARIABLES

5.1 Introduction

Chapters 5 and 6 focus on poverty and inequality. This chapter begins by reviewing the poverty and inequality concepts and measurements in Section 5.2, with particular focus on the money-metric approach. Before the poverty and inequality analyses can be conducted, the per capita income and expenditure variables need to be derived. However, as discussed in Chapter 3, some surveys (the two censuses, CS 2007, OHSs, LFSs and GHSs) contained households with reported zero or unspecified income or expenditure, and excluding them from analyses would result in unreliable poverty and inequality estimates and trends. Hence, the SRMI approach as discussed in Section 3.8.4 is used to impute the income or expenditure of these households, and this is discussed in Section 5.3. Section 5.4 explains the derivation of the per capita variables in 2000 prices. In Section 5.5, the income and expenditure variables are compared with the national accounts current income (also in 2000 prices) to investigate if some surveys seriously under-reported income or expenditure, and whether the possible under-estimation has to do with the factors discussed in Chapter 3. Section 5.6 concludes the chapter. The per capita variables derived in this chapter will be used in Chapter 6 when poverty and inequality trends are looked at.

5.2 A review of poverty and inequality concepts and measurements

5.2.1 Poverty

In this section, the concept of poverty and its dimensions are explained, before the measurement of poverty is discussed.

5.2.1.1 Definitions

Poverty, defined in general as deprivation in well-being (World Bank 2000), exists in a society “when one or more persons do not attain a level of material well-being deemed to constitute a reasonable minimum by the standards of that society” (Ravallion 1992). This definition implies the following (Boltvinik 2001): (1) Poverty and the poor are associated with a state of want, with deprivation; (2) Such deprivation is related to the necessities of life;

(3) Poverty involves a comparison between the observed and normative (or standard) condition.

With regard to the concept 'well-being' (and hence poverty), there are three approaches to it (Sen 1987; Haughton and Khandker 2009: 2-3):

- It could be thought of as the command over commodities in general. That is, people are better off (non-poor) if they have a greater command over resources to meet their needs. In order for this to happen, the individuals must have sufficient income or consumption. Hence, poverty is measured by comparing individuals' income with some defined threshold, below which they are regarded as poor.
- The second approach is to determine if people are able to obtain enough of a specific type of consumption good, such as food, shelter, health care, education, etc. Thus, this approach goes beyond the conventional monetary measures of poverty (as in the first approach). For instance, nutritional poverty could be measured by examining whether children are stunted or wasted, and educational poverty could be measured by investigating whether people are literate or the years of formal education they attained.
- The third and broadest approach argues that poverty takes place when people lack key capabilities to function in society, and hence have inadequate income, low educational attainment, poor health, feel insecure and/or a sense of powerlessness, have low self-confidence, or do not have rights such as freedom of speech. Viewed in this way, poverty is a multi-dimensional concept.

As far as the multi-dimensional nature of poverty is concerned, Chambers (1988) and the World Bank (2000: 16-21) argue that the following dimensions are involved:

- *Poverty proper*: Lack of adequate income, as well as lack of physical assets, household equipments and facilities such as formal dwelling, electricity, piped water, telecommunication equipments like landline telephone and cellphone, toilet facility in dwelling, and frequent refuse removal.
- *Health and education deprivation*: Population characterised by high death rate and infant mortality rate, as well as high incidence of under-nutrition, sickness or disability. This implies an inferior level of physical well-being of the population, and these people are defined as poor. In addition, if indicators like the net primary enrolment rate are low, then it is highly likely that the majority of the population is poor, as the population fails to possess the human capital required to generate the income level needed to attain a reasonable level of material well-being.

- *Physical or social isolation*: This happens due to reasons like peripheral location, lack of access to goods and services, illiteracy, ignorance, discrimination or stigma.
- *Voicelessness and powerlessness*: It means the lack of links to networks (e.g., family-based networks, occupation-based groups of mutual help, savings and credit groups, etc.), and could happen due to physical or social isolation as discussed above. Criteria such as trust, accountability, participation, ability to build unity, respect, responsiveness, fairness and caring, as well as loving and listening, could be used to measure voice and power.
- *Vulnerability*: This means the possible exposure to crisis and the risk of becoming even poorer. For example, the risk that an individual will experience an event of income or health poverty over time, and the likelihood of being exposed to a number of other risks, such as crime, violence, natural disasters, being pulled out of school. Households or individuals with insufficient physical assets[80], human capital[81], links to networks[82], as well as income diversifications[83] are more vulnerable to the abovementioned risks.

Hence, from the discussion above, it is obvious that poverty has both money-metric dimension and non-money-metric dimension.

Boltvinik (2001) argues that "a comparison between an observed and a normative condition" is required to determine the extent of poverty, and this comparison can be made objectively or subjectively. Objective comparisons are, in general, associated with quantitative measures, and define poverty as economic deprivation (in terms of income, expenditure / consumption or asset possession), educational deprivation (inferior level of education) and biological deprivation (malnutrition, chronic disease or a disabling condition). In contrast, subjective comparisons are generally linked with qualitative measures and identify poverty as physical or social isolation (due to reasons like peripheral location of residence, lack of access to goods and services, etc.), powerlessness within existing social, economic, political and cultural structures, vulnerability to a crisis, the risk of becoming poorer, lack of work, or being

[80] Individuals with physical assets are less vulnerable because they have the capacity to self-insure by selling the physical assets to compensate for temporary loss of income (World Bank 2000: 20).
[81] Individuals with limited education are more vulnerable to income fluctuations and less able to manage risk due to, for example, lack of access to credit or multiple income sources (World Bank 2000: 20).
[82] Networks like occupation-based groups of mutual help, rotating savings and credit groups and family-based networks could reduce a person's vulnerability by providing transfers in cash or kind in the event of calamity (World Bank 2000: 20).
[83] It is often argued that people in urban areas are less vulnerable, because their non-farm income, which comes from various sources like non-agricultural employment, investment, etc., comprises the majority of their income. Farm income fluctuates more than non-farm income, thus failing to provide a measure of protection against weather-related risks (World Bank 2000: 20).

involved in arduous, hazardous work.

Poverty is not a static condition but includes the time dimension, as it is possible for a poor person to rise out of poverty or an affluent person to suffer a financial or non-financial reversal (Carter and May 2001). Hence, temporary and chronic poverty could be distinguished: temporary poverty means some people move between being poor and non-poor over time. This may be a result of one-time decline in living standards (e.g., loss of a job, natural disasters, seasonal variations in food security). In contrast, chronic poverty means some people are continuously poor. It is more difficult to address chronic poverty, as it is often associated with persistent inter-generational poverty (Woolard 2001: 98).

Having defined poverty, the next step is to measure poverty, and three steps need to be taken (Haughton and Khandker 2009: 10):

- Defining a welfare indicator;
- Determining the poverty line, i.e., a minimum acceptable standard of the indicator so as to distinguish the poor from the non-poor;
- Finding out how much poverty there is by generating a summary statistic, so as to aggregate information from the distribution of the indicator relative to the poverty line.

Each of these steps is discussed in detail below.

5.2.1.2 Welfare indicator

With regard to the poverty measurement that only takes non-money-metric welfare indicators(s) into account, there are two general approaches. In the first approach, a basic need is specified (e.g., having a formal dwelling, access to water inside dwelling, electricity as an energy source for cooking, household head being able to read and write, etc.) as the welfare indicator, and the poor is defined as anybody who is deprived in this dimension. A drawback of this approach is that it only takes one indicator into consideration at a time, and fails to estimate tradeoffs among the dimensions (World Bank 1990).

For instance, assume person A stays in a formal dwelling but does not have access to electricity as fuel source for cooking, person B has electricity but resides in an informal dwelling, and person C does not have electricity and resides in an informal dwelling. If dwelling type is used as the welfare indicator, and residing in a formal dwelling is the minimum acceptable standard of the indicator, persons B and C are defined as poor. In contrast, if fuel source for cooking is the welfare indicator, and having access to electricity is the minimum acceptable standard, persons A and C are identified as poor. From the results above, it is obvious that C is the poorest of the three, but it is difficult to determine whether person A or person B is poorer.

Hence, the second approach is adopted. In this approach, numerous non-money-metric variables are synthesized into a composite welfare index by a statistical procedure. The result, the poverty, marginality or socio-economic status (SES) index is in the form of a number without specific content, which is subsequently used to rank geographical areas (e.g., province) or demographic groups (e.g., race) from the more deprived to the less so (Boltvinik 2001: 11-13).

Equal weights or different weights could be allocated to each variable. An example of this approach can be found in the second Southern Africa Consortium for Monitoring Educational Quality (SACMEQ II) study on Grade 6 pupil literacy and mathematics performance (Moloi and Strauss 2005). Question 7 of the pupil questionnaire asked the pupil participants to declare if they have got certain assets or facilities available at the place of residence[84], and the answers from these questions were used to derive the SES index, which has a minimum score of 0 (if the answer is "no" to all questions) but a maximum score of 14 (if the answer is "yes" to all questions). Those with an SES index below seven and at least seven are defined as people with low SES status (i.e., poor) and high status (i.e., non-poor) respectively. Another example is that a composite welfare index could be constructed using seven non-money-metric variables in CS 2007 (the index has a minimum score of 0 but a maximum score of 7), as presented in Table 5.1.

However, attaching equal weights to the variables as adopted in Table 5.1 might not be the best approach. For example, if 95% of the sample has access to electricity in their dwellings but only 50% has landline telephones, then it could be argued that the former variable should be given greater weight, because a very high proportion of people have access to electricity, so the remaining 5% who do not have access to electricity feel very inferior and the lack of access to electricity should clearly indicate these people are poor.

[84] There are 14 variables in total: daily newspaper, weekly or monthly magazine, TV set, radio, video cassette recorder, cassette player, telephone, refrigerator/freezer, car, motorcycle, bicycle, electricity, piped water and a table to write on.

Table 5.1: Welfare indicators used to derive the poverty index in CS 2007

Indicator	Index = 1	Index = 0
Dwelling	House or brick structure Flat in a block of flats Town / Cluster / Semi-detached house	Other
Fuel source for cooking	Electricity Solar	Other
Water	Piped water inside dwelling Piped water inside yard	Other
Sanitation	Flush toilet (connected to sewerage system) Flush toilet (with septic tank) Chemical toilet	Other
Refuse removal	Removed by local authority once a week	Other
Telephone in dwelling or cellphone	Yes	No
Employment status of household head	Employed	Other
Education attainment of household head	At least Matric	Other

The principal component analysis (PCA) method is often adopted to deal with this problem. This technique attaches the most weight to the asset variables that are most unequally distributed, i.e., the greater the standard deviation of a variable, the greater its weight. The range of variables is analysed so as to extract those linear combinations of the variables that capture the most common information. Each linear combination or "principal component" is uncorrelated with the others, in order to capture a different dimension in the data. The first principal component explains the most variation in the data, with successive components explaining additional but less variation. The first principal component is commonly used for the construction of the SES index. Once the SES index is derived, a relative poverty line (e.g., the SES index that distinguishes the poorest 40%) is chosen to analyse the characteristics of the poor. Examples of recent South African studies adopting this approach are Bhorat et al. (2006) and Bhorat, Van der Westhuizen and Goga (2007)).

It is also possible to derive a welfare index that includes both money-metric and non-money-metric variables and by methods like PCA or a fuzzy sets approach in order to measure poverty. A good example is the Human Development Index (HDI) derived by the United Nations Development Program (UNDP). It measures a country's welfare using the following three dimensions of human development (Bhorat et al. 2004; Govender et al. 2006): (1) a long and healthy life, as measured by life expectancy at birth index; (2) knowledge, which is measured by an education index that evaluates both adult literacy and the general enrolment in primary, secondary or tertiary education; (3) a decent standard of living, as measured by the gross domestic product (GDP) per capita index. Equal weight is allocated to each dimension.

Another example is the fuzzy sets approach fuzzy sets approach adopted by Burger, Van der Berg, Van der Walt and Yu (2004) that takes household income and numerous non-money-metric household characteristics (e.g., dwelling type, source of water, sanitation, etc.) into consideration. However, the abovementioned approaches fall beyond the scope of this dissertation and will not be discussed in detail.

The discussions above relate to the so-called "direct approach". An alternative to it is the indirect or income approach, which first measures the resources (which involve not only income, but entitlement or rights) that a household commands, before comparing the magnitude and composition of these resources with the resource requirement to meet the set of basic needs. If the resources identified are reduced to income or expenditure, the methodology is referred to as poverty line, because the welfare indicator is expressed as a quantity of money (i.e., money-metric variable) (Boltvinik 2001). Two commonly used methods are to work out the cost of a minimum basket of goods and then use the required income or expenditure level as the poverty line, and to estimate the income or expenditure that allows an individual to obtain food to meet energy requirements for survival. They will be discussed in detail when poverty lines are dealt with.

5.2.1.3 Poverty line

A poverty line divides the population into two groups on the basis of some measure: below the line a person is considered as poor, and above the line he/she is classified as non-poor. By defining a line that is regarded as some kind of minimum level (to be discussed in detail below), one is able to find out the number of poor people, as well as the depth and severity of poverty. However, a poverty line will always be an imperfect construct, as the point at which the line is drawn is somewhat random and often highly contentious (Woolard 2001: 93). For instance, at an income poverty line of R500, a person earning R499 is in poverty, while another person earning R501 is not. These two persons' poverty statuses are different, despite the income difference of R1. Yet, for purposes of understanding the nature of poverty, it is necessary to draw the poverty line. This sub-section will discuss the difference between absolute and relative poverty lines, and objective and subjective poverty lines, as well as the impact of household size and structure on the poverty line.

<u>Absolute poverty line vs. Relative poverty line</u>

Absolute and relative poverty lines are distinguished in the literature. The <u>absolute poverty line</u> is not meant to change with the society's standard of living. Instead, it is an objective, scientific determination, because it is based on the minimum requirement needed to sustain life, and is usually based on essential goods and nutritional needs (Woolard and Leibbrandt 2000:46; Govender et al. 2006). In addition, if economic growth takes place, it would result in a reduction in the number of people in absolute poverty (Woolard and Leibbrandt 2006: 18).

In South Africa, the three absolute poverty lines (per capita per month, 2000 prices) proposed by Woolard and Leibbrandt (2006) have been used in some recent studies for deriving the poverty estimates and trends[85]. These lines were derived as follows:

- Food poverty line (R211): the expenditure needed to purchase enough food to meet the basic daily food-energy requirement of 2 261 kilocalories for the average person over one month, as recommended by the South African Medical Research Council (MRC);
- Lower bound poverty line (R322): this was calculated by observing the essential non-food expenditure of households that spent approximately R211 on food. It was found that these households spent R111 on essential non-food items. Hence, the lower bound poverty line was equal to R322 (= R211 + R111).
- Upper bound poverty line (R593): this was derived by observing the total non-food expenditure of households that spent roughly R211 on food. It was found that these households spent R382 on all non-food items. Therefore, the upper bound poverty line was equal to R593 (= R211 + R382). This also implies the expenditure on non-essential non-food items of the abovementioned households was R271 (= R382 – R111).

In contrast, a <u>relative poverty line</u> moves with standards of living, and the poor are taken to be those individuals who are suffering from relative deprivation (Woolard and Leibbrandt 2000: 47). There are two interpretations of those classified as relatively poor (Boltvinik 2001; Govender et al. 2006; Woolard and Leibbrandt 2006; Haughton and Khandker 2009):

- The poorest x% (e.g., 40%) of the population is poor. In other words, the poverty line defines the poor as the population in certain specified deciles. This implies that the focus is on the poorest segment of the population, and "the poor are always with us" (Haughton and Khandker 2009: 43), as the percentage distinguished as poor remains the same, regardless of whether their circumstances have improved.
- The poor are defined as such if their living standard, as measured by income or expenditure, is below a percentage of that of their contemporaries, for instance, 50% of mean income or expenditure). In this way, the percentage of poor is not preset (as

[85] These three poverty lines will also be used when looking at poverty trends since the transition using different household surveys, and the results will be discussed in Chapter 6.

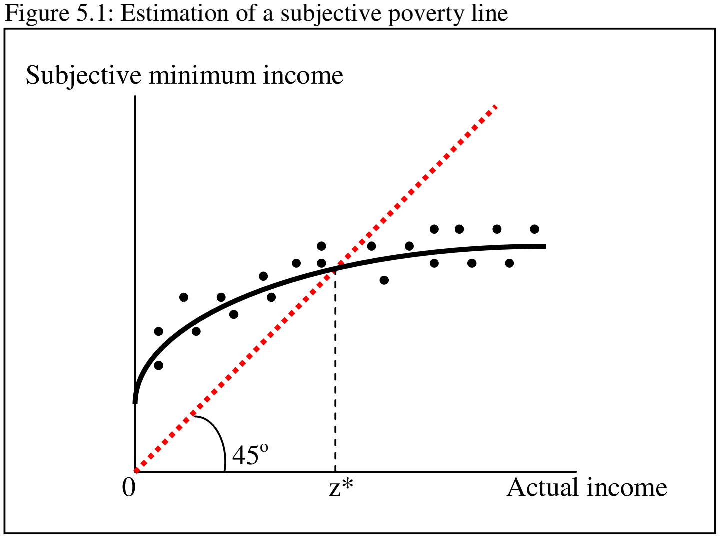
compared with the first interpretation), but neither is the level of this standard of living. Furthermore, only a change in the distribution of income will decrease the number of people in relative poverty.

Objective poverty line vs. Subjective poverty line

Since the notion of basic needs, including nutritional requirements, varies from person to person, a subjective poverty line is estimated. In other words, as people make subjective and different judgments regarding what constitutes a socially acceptable minimum standard of living in a society, different poverty lines are used by the individuals (Ravallion 1992: 33 & 1998: 21).

An example is the answer to a question such as "what income level do you personally consider to be absolutely minimal? That is to say, with less you could not make ends meet". The answer differs amongst the respondents, but it tends to be an increasing function of actual income. These answers could be plotted, with a line fitting through them, before a subjective poverty line, as indicated by z* in Figure 5.1, is derived. People with income below z* tend to feel their income is inadequate (i.e., actual income is lower than minimum income considered to be sufficient for survival), while people with income above z* feel their income is adequate.

Figure 5.1: Estimation of a subjective poverty line



Source: Haughton and Khandker (2009: 61).

Compared with the subjective poverty line, an objective poverty line is estimated by looking at the nutritional requirements for a healthy and active life (Ravallion 1998: 8), with the two commonly used approaches being the cost of basic needs method and the food energy intake method. The cost of basic needs (CBN) method works as follows (World Bank 1990;

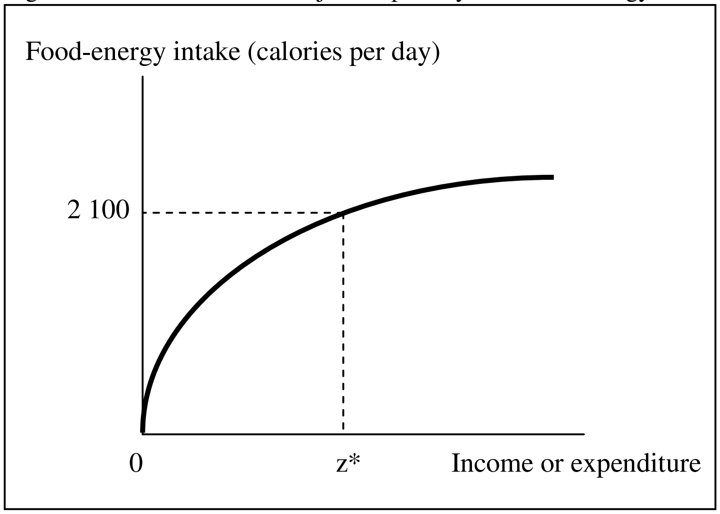
Ravallion 1992: 26-27 & 1998:15; Woolard 2001: 94; Woolard and Leibbrandt 2006: 21; Haughton and Khandker 2009: 49-50):

- A consumption bundle, with both food and non-food components, that is deemed to be adequate for meeting basic needs, is stipulated.
  - The consumption of the food items should result in meeting the minimum number of calories (such as 2 100 calories per person per day) required for good health.
  - Non-food necessities include items like education, health, sanitation, water supply, housing and telecommunication.
- The prices of the food and non-food items are collected.
- The minimum expenditure required to obtain these food and non-food necessities is estimated, and this amount forms the poverty line. In other words, the basic needs poverty line (z^{BN}) is equal to the sum of the food component (z^{F}) and non-food component (z^{NF}), or $z^{BN} = z^{F} + z^{NF}$.

The CBN method is not without shortcomings (World Bank 1990; Ravallion 1992: 26-27 & 1998: 17; Woolard 2001: 94-95; Haughton and Khandker 2009: 50). First, this poverty line might need to be adjusted for different areas (e.g., urban vs. rural areas) within the country, because prices or access to goods and services could differ. Secondly, price data might not be available for all goods in the consumption bundle. Thirdly, it is subjective to say which non-food components are necessary for meeting basic needs. For example, steak could be regarded as a necessity by relatively more privileged people residing in urban areas, but a luxury by those residing in rural areas. Thirdly, with regard to the food components, the number of calories required for good health is highly variable from one person to another, depending on their metabolism and activity levels. Even if it is the same for all individuals (e.g., 2 100 calories, as discussed above), many bundles of food items yield the same nutrition, and the cost of each bundle differs.

When price data are not available, the food energy intake (FEI) method is the alternative to construct the poverty line. The method works a follows (Ravallion 1998: 10; Haughton and Khandker 2009: 54-55): First, the income or expenditure level at which food energy intake is sufficient to meet pre-determined food energy requirements is found. Next, a calorie income function is estimated. In other words, the vertical axis is food-energy intake, and it is plotted against total income (or expenditure) on the horizontal axis (Figure 5.2). A line of best fit is indicated, that is, the expected value of caloric intake at a given value of income (or expenditure). It is then possible to find the expenditure z* at which a person normally attains the stipulated food-energy requirement.

Figure 5.2: Estimation of an objective poverty line: Food energy intake method



Source: Ravallion (1998: 11).

A merit of this method is that it avoids the need for price data. In addition, it automatically includes an allowance for both food and non-food expenditure, as long as the total expenditure at which the person attains the required calorie level is located (Ravallion 1998: 11). However, the method also involves drawbacks (Ravallion 1992: 28 & 1998: 11; Haughton and Khandker 2009: 57-59); it is not easy to set the food-energy requirement, as the relationship between food energy intake and income (or expenditure) is not the same across regions, sectors or dates, but will shift according to differences in relative prices, tastes, etc. Hence, as in the CBN method, there might be a need to establish different poverty lines, for example, by area type and by year.

Adjusting for household size and structure

As discussed previously, the commonly used money-metric variable for measuring poverty is per capita income or expenditure, which is derived by dividing household income (or expenditure) by household size. However, strictly speaking, households differ not only in size but also by demographic make-up, and hence the straightforward comparison of per capita income (or expenditure) might be deceptive (Woolard and Leibbrandt 1999: 12). For instance, the cost of a child in terms of food expenditure required for survival is smaller than that of an adult[86]; there are economics scale involved in consumption for items like housing, since it costs less to house a couple than to house two individuals separately. Hence, per adult

[86] There are also equivalence scales that assign an adult male equivalence less than one to adult females, as adult females tend to consume less of most goods than adult males do (Ravallion 1992: 17-18).

equivalent income or expenditure should be used for measuring poverty. In other words, adult equivalence scale (AES) is involved.

There exists a wide range of AES, with the simplest and most commonly used being the so-called 'double parameter class of scales' introduced by Cutler and Katz (1992), namely $E = (A + \alpha C)^{\theta}$, where:

E = number of adult equivalents in the household

A = number of adults in the household

C = number of children in the household

α = a constant reflecting the resource cost of a child relative to an adult, with $0 \leq \alpha \leq 1$

θ = the overall economies of scale within the household, with $0 \leq \theta \leq 1$

If α is equal to one, it means children are counted as adults. θ has the same range, and if it is smaller than one, economies of scale in consumption is taken into consideration (Woolard 2001; Haughton and Khandker 2009).

There are no universal, scientifically determined correct values of these two parameters. With regard to the derivation of the value of α, a common approach is to compare the energy requirements for adults and children. However, since the two groups consume food and non-food items, there is no reason to expect non-food costs to follow the same ratio (Streak, Yu and Van der Berg 2009: 186). An alternative method to derive α is the Engel method[87].

In addition to the scale proposed by Cutler and Katz, there are other AESs. First, with regard to the square root scale, it approximates the number of adult equivalents as the square root of household size in order to address the economies of scale. This scale does not distinguish the different needs of adults versus children (i.e., α equals 1).

Looking at the two OECD scales, the original scale assigns a value of one to the first adult household member, of 0.7 to each additional adult and of 0.5 to each child. Thus, the 0.7 reflects economies of scale; the smaller this parameter, the higher the importance given to economies of scale. Moreover, the 0.5 is the weight given to children, and this reflects their lower needs (e.g., food, housing space, etc.) (Haughton and Khandker 2009: 29). The OECD scale was later modified by assigning a value of 0.5 (instead of 0.7) to each additional adult

[87] However, this falls beyond the scope of the study. Refer to Woolard (2001) for detail.

and of 0.3 (instead of 0.5) to each child. Table 5.2 summarizes the abovementioned equivalence scales.

Table 5.2: Commonly used adult equivalence scales

Scale	Equation
Square root scale	$E = (A+C)^{0.5}$
OECD original	$E = 1 + 0.7(A-1) + 0.5C$
OECD modified	$E = 1 + 0.5(A-1) + 0.3C$
Double parameter class of scales	$E = (A+\alpha C)^{\theta}, \quad 0 < \alpha \leq 1, \quad 0 < \theta \leq 1$

Sources: Deaton and Paxton (1997); OECD (2008); Streak, Yu and Van der Berg (2009).

Some recent studies showed that the choice of equivalence scale made a small difference to the identification of the poor; and the poverty measures did not differ too much (although it does not necessarily mean that the same group of people are distinguished as poor at different scale parameters), regardless of whether per capita or per equivalent variables were used (e.g., May, Carter and Posel 1995; Woolard and Leibbrandt 2000; Woolard 2001; Streak et al. 2009). For instance, Woolard and Leibbrandt (2000) applied various Cutler and Katz scales[88] on the IES 1995 data. Fixing the share of households in poverty at 40%, it was found that the poverty profile changed very little even when quite large adjustments were made to the scale parameters. In particular, the poverty rate amongst blacks, coloured, as well as urban and rural dwellers remained unchanged. Streak et al. (2009) conducted a similar study on IES 2005/2006, but focusing on child[89] poverty, and found that the magnitude and composition of child poverty was not sensitive to the scale used. Hence, per capita income/expenditure/consumption variables rather than per adult equivalent variables would be derived and used for the poverty and inequality analyses in Chapter 6.

5.2.1.4 Measurement

Having identified the indicator of welfare and establishing the poverty line, the next step is to use the household survey data to measure poverty[90]. The following four axioms form the basis of what has become a widely accepted consensus with regard to the basic requirements of a good poverty measure (Sen 1976):

- *Monotonicity*: The poverty index must rise (fall) if the income of a poor person decreases (increases).
- *Transfer*: If a poor individual transfers his/her income to someone less poor than

[88] The following scales were used: (1) $\alpha = 0.5, \theta = 0.6$; (2) $\alpha = 0.5, \theta = 0.75$; (3) $\alpha = 0.75, \theta = 0.6$; (4) $\alpha = 0.75, \theta = 0.75$; (5) $\alpha = 0.75, \theta = 0.9$; (6) $\alpha = 1, \theta = 0.6$; (7) $\alpha = 1, \theta = 0.75$; (8) $\alpha = 1, \theta = 0.9$.

[89] Children were defined as those aged 0-17 years in their study.

[90] A limitation of using household survey data to measure poverty is that one distinct sub-group of the poor is missed: those who are homeless (Ravallion 1992).

himself/herself (regardless of whether the latter person is poor or non-poor), the index must rise.

- *Population symmetry*: The index must not change if two or more identical populations are pooled.
- *Proportion of poor*: The index must rise (fall) if the proportion of the population defined as poor increases (decreases).

The most commonly used poverty measures are those proposed by Foster, Greer and Thorbecke (1984)[91], which could be expressed as follows:

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^\alpha \Bigg| (y_i \le z)$$

Where

P_α = measure of poverty

q = number of poor people

n = total number of people

z = poverty line

y_i = income of the i-th person in the population

Three poverty measures could be derived from the equation above, and these measures are explained below with the aid of the information from Table 5.3.

Table 5.3: Income of each person in a hypothetical country, 2000-2002

	Poverty line: R3 000				
	Income for each individual in the country (Rand)				
n = 5	**Person A**	**Person B**	**Person C**	**Person D**	**Person E**
2000	5 000	5 000	2 500	2 500	2 500
2001	5 000	5 000	2 000	2 000	1 000
2002	5 000	5 000	2 250	2 250	500

<u>Headcount index</u>

The headcount index (H) is simply the proportion of the population that is poor, i.e., $H = P_0 = \frac{q}{n}$. Although the index is very easy to interpret by showing the <u>incidence</u> of poverty, it ignores the extent of which the poor fall below the poverty line, because it fails to indicate how poor the poor are. Hence, the index does not change even if people below the poverty line become poorer (World Bank 1990:27; Haughton and Khandker 2009: 69). For

[91] For the remainder of the dissertation, it will be referred to as the FGT poverty measures.

example, Table 5.4 indicates that the headcount index is the same (0.6) in all three countries, but comparing the three poor persons between 2000 and 2001, they were clearly poorer in 2001, as their income decreased. Hence, poverty became more serious in 2001, yet the headcount index remains unchanged. In other words, the *monotonicity axiom* is not met.

Table 5.4: Poverty headcount index of a hypothetical country, 2000-2002

	Poverty line: R3 000				
	Income for each individual in the country (Rand)				
n = 5	**Person A**	**Person B**	**Person C**	**Person D**	**Person E**
2000	5 000	5 000	2 500	2 500	2 500
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_0 = \frac{3}{5} = 0.6$				
2001	5 000	5 000	2 000	2 000	1 000
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_0 = \frac{3}{5} = 0.6$				
2002	5 000	5 000	2 250	2 250	500
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_0 = \frac{3}{5} = 0.6$				

Furthermore, the headcount index does not meet the *transfer axiom*, because a transfer from a poor person to someone less poor does not result in an increase in the index (Govender et al. 2006: 15-16)[92], thereby failing to capture the severity of poverty. For example, using the information in Table 5.4 again, between 2001 and 2002, person E, the poorest person, transferred R250 to person C and another R250 to person D, with both person C and D being less poor than person E before the transfer took place. However, after the transfer, the poverty index remains unchanged at 0.6, despite the fact that the severity of poverty of person E increased between the two years, as his income was further below the poverty line.

Although the poverty headcount index fails to meet the two abovementioned axioms, it does meet the *population symmetry* and *proportion of poor* axioms. If the 2000-2002 information from Table 5.3 above is pooled, the poverty headcount index remains at 0.6 (i.e., the population symmetry axiom is met). On the other hand, if the income of person B in 2000 had decreased from R5 000 to R2 500, this means the number of poor would have increased from three to four persons, and consequently the headcount index increased from 0.6 to 0.8 (i.e., the proportion of poor axiom is met).

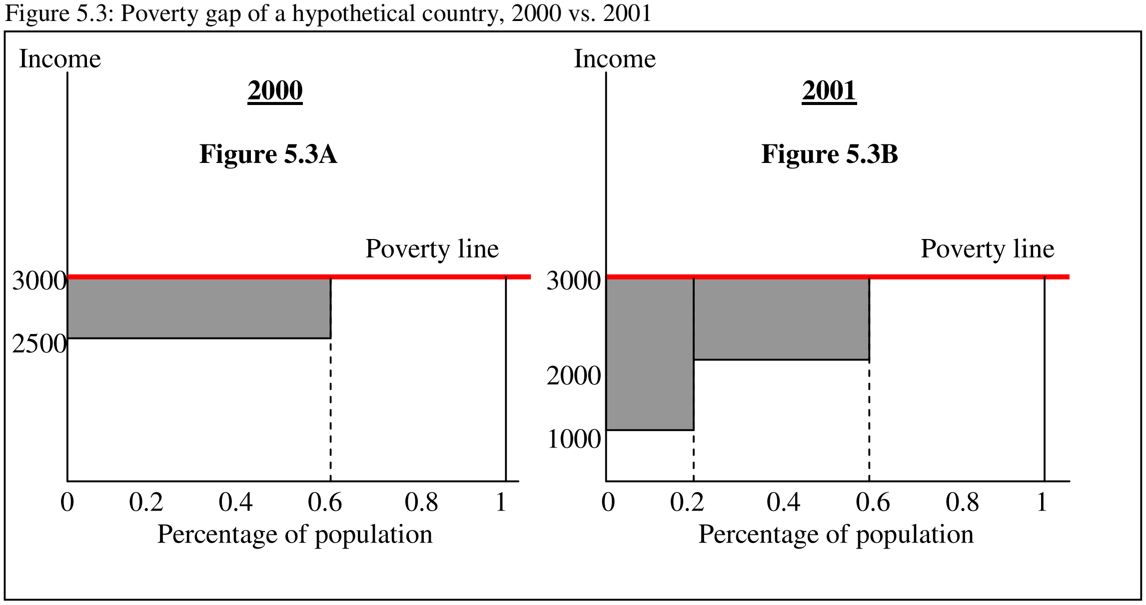
[92] In fact, the index could fall if the net distribution from the very poor to the just-poor results in the latter group of people being lifted out of poverty.

Poverty gap index

Poverty gap index (PG) adds up the extent to which the poor on average fall below the poverty line. (Haughton and Khandker 2009: 70), and is equal to the total amount of income necessary to raise everyone who is below the poverty line up to that line (i.e., poverty gap), as a proportion of the total income of the population if their income had been equivalent to the poverty line, i.e., $PG = P_1 = \frac{1}{n}\sum_{i=1}^{q}\frac{z - y_i}{z}$.

Before the index can be derived, incomes need to be arranged in ascending order, with the poorest earning y_1, the next poorest y_2, and so forth, with the least poor earning y_q (Ravallion 1992). Next, the poverty gap is calculated, and it depends on the distances of the poor below the poverty line. For example, Figure 5.3 below provides a graphical illustration of what happened in 2000 and 2001 using the information from Table 5.3. It was shown earlier that 60% of the people were poor in both years. However, the shaded area in Figure 5.3B is larger, implying that a greater amount of income is needed to eliminate absolute poverty in 2001. Hence, the poverty gap and subsequently poverty gap index would be greater in 2001.

Figure 5.3: Poverty gap of a hypothetical country, 2000 vs. 2001



The shaded areas in the two figures above assume that the policy makers know who the poor are and how much they currently earn, and the total cost involved to eliminate poverty by targeting transfers to the poor is equivalent to $\sum_{i=1}^{q} z - y_i$. That is, total cost is R1 500 (R500 + R500 + R500) in 2000 but R4 000 (R2 000 + R1 000 + R1 000) in 2001. However, if the policy maker does not know who is poor and who is not, he/she would have to give an amount equivalent to the poverty line so as to be sure that none are poor. In other words, the cost is equal to zn. This means in both 2000 and 2001, the total cost involved to eliminate poverty would have been equal to R15 000 (R3 000 × 5).

The detailed derivation of the poverty gap index in each year (using the information from Table 5.3) is illustrated in Table 5.5. The *monotonicity axiom* is met, as indicated by the indices in 2000 and 2001, because the poverty gap increased as the income of persons C, D and E decreased between the two years. That is, the depth of poverty is reflected by this index. It also meets the *population symmetry axiom*, because pooling two or more identical populations results in no change in the index.

Table 5.5: Poverty gap index of a hypothetical country, 2000-2002

	Poverty line: R3 000				
	Income for each individual in the country (Rand)				
n = 5	**Person A**	**Person B**	**Person C**	**Person D**	**Person E**
2000	5 000	5 000	2 500	2 500	2 500
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_1 = \frac{1}{5}\left[0+0+\left(\frac{3000-2500}{3000}\right)+\left(\frac{3000-2500}{3000}\right)+\left(\frac{3000-2500}{3000}\right)\right] = \frac{1}{5}\left[\frac{1500}{3000}\right] = 0.10$				
2001	5 000	5 000	2 000	2 000	1 000
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_1 = \frac{1}{5}\left[0+0+\left(\frac{3000-2000}{3000}\right)+\left(\frac{3000-2000}{3000}\right)+\left(\frac{3000-1000}{3000}\right)\right] = \frac{1}{5}\left[\frac{4000}{3000}\right] = 0.27$				
2002	5 000	5 000	2 250	2 250	500
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_1 = \frac{1}{5}\left[0+0+\left(\frac{3000-2250}{3000}\right)+\left(\frac{3000-2250}{3000}\right)+\left(\frac{3000-500}{3000}\right)\right] = \frac{1}{5}\left[\frac{4000}{3000}\right] = 0.27$				

However, the index fails to meet the *transfer axiom*. Looking at Table 5.5 again, the index was unchanged at 0.27 in 2001 and 2002, despite the fact that person E (the poorest) transferred his income to persons C and D (the next two poorest persons). Hence, as in the poverty headcount index, poverty gap index is insensitive to the extent of inequality among the poor (World Bank 1990).

The only exception is that if the transfers from the poor to the people less poor result in the latter becoming non-poor, the poverty gap index would increase (Woolard and Leibbrandt 1999: 20-21). This could be explained by the information in Table 5.6. In 2007, the poverty gap index was 0.16. Assuming in 2008 (Case #1), person J (the poorest) transferred his income to persons H and I (R150 to each), but the incomes of the latter persons remained below the poverty line after the transfer. Hence, there was no change in the poverty gap index. In contrast, had person J transferred a bigger amount of his income to these two people (R300 to each, i.e., Case #2) to enable them to be out of poverty (their income increased to R3 100), the poverty gap index would have increased from 0.16 to 0.17.

Table 5.6: Poverty gap index of a hypothetical country, 2007-2008

<table>
<tr><td rowspan="2"></td><td colspan="5">Poverty line: R3 000</td></tr>
<tr><td colspan="5">Income for each individual in the country (Rand)</td></tr>
<tr><td>n = 5</td><td>Person F</td><td>Person G</td><td>Person H</td><td>Person I</td><td>Person J</td></tr>
<tr><td rowspan="3">2007</td><td>5 000</td><td>5 000</td><td>2 800</td><td>2 800</td><td>1 000</td></tr>
<tr><td>Non-poor</td><td>Non-poor</td><td>Poor</td><td>Poor</td><td>Poor</td></tr>
<tr><td colspan="5">$P_0 = \frac{3}{5} = 0.6$
$P_1 = \frac{1}{5}\left[0+0+\left(\frac{3000-2800}{3000}\right)+\left(\frac{3000-2800}{3000}\right)+\left(\frac{3000-1000}{3000}\right)\right] = \frac{1}{5}\left[\frac{2400}{3000}\right] = 0.16$</td></tr>
<tr><td rowspan="3">2008 (Case #1)</td><td>5 000</td><td>5 000</td><td>2 950</td><td>2 950</td><td>700</td></tr>
<tr><td>Non-poor</td><td>Non-poor</td><td>Poor</td><td>Poor</td><td>Poor</td></tr>
<tr><td colspan="5">$P_0 = \frac{3}{5} = 0.6$
$P_1 = \frac{1}{5}\left[0+0+\left(\frac{3000-2950}{3000}\right)+\left(\frac{3000-2950}{3000}\right)+\left(\frac{3000-700}{3000}\right)\right] = \frac{1}{5}\left[\frac{2400}{3000}\right] = 0.16$</td></tr>
<tr><td rowspan="3">2008 (Case #2)</td><td>5 000</td><td>5 000</td><td>3 100</td><td>3 100</td><td>400</td></tr>
<tr><td>Non-poor</td><td>Non-poor</td><td>Poor</td><td>Non-poor</td><td>Poor</td></tr>
<tr><td colspan="5">$P_0 = \frac{1}{5} = 0.2$
$P_1 = \frac{1}{5}\left[0+0+0+0+\left(\frac{3000-400}{3000}\right)\right] = \frac{1}{5}\left[\frac{2600}{3000}\right] = 0.17$</td></tr>
</table>

The poverty gap index also does not meet the *proportion of poor* axiom. Comparing 2007 and Case #2 of 2008 again in Table 5.6, it is obvious that, although the proportion of population defined as poor decreased (from 0.6 to 0.2), the poverty gap did not decrease. In fact, it increased from 0.16 to 0.17.

Squared poverty gap index

The squared poverty gap index, $P_2\left(=\frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^2\right)$, measures the severity of poverty by taking both poverty and inequality amongst the poor into account. More weight is put on observations that fall well below the poverty line. For example, the previous discussion relating to Tables 5.4 and 5.5 showed poverty headcount and poverty gap indices were the same in 2001 and 2002. However, if person E transferred R500 of his income to the other two poor (D and E) in 2002, person E's new income level (R500) after the transfer would end up well below the poverty line (R3 000) but persons C and D's new income level (R2 250) would be closer to the poverty line. As a result, poverty is more severe in 2002, as indicated by the greater squared poverty gap ratio in Table 5.7 (0.1639, compared with 0.1333 in 2001). Hence, this index meets the *transfer axiom*.

Table 5.7: Squared poverty gap index of a hypothetical country, 2000-2002

	Poverty line: R3 000				
	Income for each individual in the country (Rand)				
n = 5	**Person A**	**Person B**	**Person C**	**Person D**	**Person E**
2000	5 000	5 000	2 500	2 500	2 500
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_2 = \frac{1}{5}\left[0+0+\left(\frac{3000-2500}{3000}\right)^2+\left(\frac{3000-2500}{3000}\right)^2+\left(\frac{3000-2500}{3000}\right)^2\right] = 0.0167$				
2001	5 000	5 000	2 000	2 000	1 000
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_2 = \frac{1}{5}\left[0+0+\left(\frac{3000-2000}{3000}\right)^2+\left(\frac{3000-2000}{3000}\right)^2+\left(\frac{3000-1000}{3000}\right)^2\right] = 0.1333$				
2002	5 000	5 000	2 250	2 250	500
	Non-poor	Non-poor	Poor	Poor	Poor
	$P_1 = \frac{1}{5}\left[0+0+\left(\frac{3000-2250}{3000}\right)^2+\left(\frac{3000-2250}{3000}\right)^2+\left(\frac{3000-500}{3000}\right)^2\right] = 0.1639$				

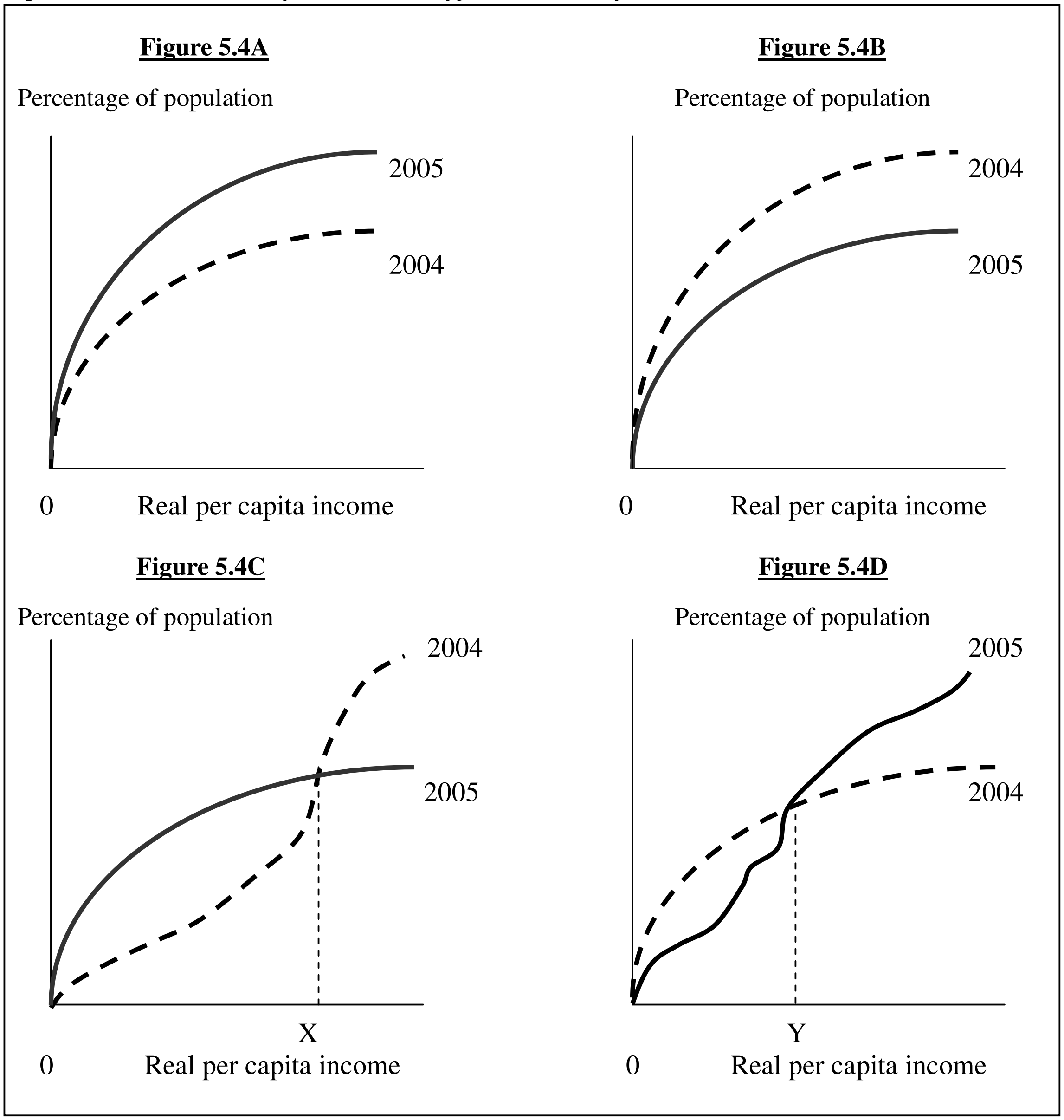
Although the squared poverty gap index has clear advantages for certain purposes, such as comparing policies which aim to reach the poorest, a major criticism of it is that it is not easy to interpret and hence is not used widely (Ravallion 1992).

In addition to the three FGT indices as discussed above, the other commonly used methods to measure poverty are cumulative density functions (CDFs) for dominance testing and growth incidence curves (GICs), and they will be discussed in detail below.

<u>Cumulative density function for dominance testing</u>

In a cumulative density function (CDF), the vertical axis shows the percentage of total population with real income that is less than or equal to the real income value on the horizontal axis. As real income increases, the corresponding cumulative proportion of population will also increase. The strength of this approach is that it makes it possible to compare the changes in poverty from one period to the next (or poverty amongst various demographic groups, such as race, gender, province), independent of any single poverty line. If the CDF for a given period lies above the CDF for the previous period on the horizontal axis, this means that poverty has increased, irrespective of any given poverty line, because the percentage of population with a certain real income or less has increased (i.e., Figure 5.4A). If the opposite happens, this means poverty has decreased at all poverty lines (i.e., Figure 5.4B).

Figure 5.4: Cumulative density functions of a hypothetical country, 2004 vs. 2005



However, if the two CDFs cross each other, this implies that comparison of poverty estimates between two periods is sensitive to the poverty line chosen. For example, in Figure 5.4C, at any income level below X, poverty in 2004 was lower. In contrast, at any income level greater than X, the 2005 CDF lies below the 2004 CDF, indicating that poverty was lower in 2005. Similarly, looking at Figure 5.4D, using any poverty line lower than Y would lead to the conclusion that poverty decreased between 2004 and 2005, but the opposite conclusion is reached if any poverty line greater than Y is used.

CDFs for dominance testing will be applied across the household surveys, when poverty trends are dealt with in Chapter 6.

5.2.2 Inequality

5.2.2.1 Definitions

Inequality, which considers the variations in the standards of living across the whole population, is a broader concept than poverty, since it is defined not just for the portion of the population below a certain poverty line, but over the entire population (Haughton and Khandker 2009). Before the inequality measurements are discussed, one must first distinguish between absolute inequality and relative inequality. The former depends on absolute differences (e.g., in Rand values) in the levels of income, while the latter is influenced by the ratios of individual incomes to the overall mean (Govender et al. 2006)[93]. Relative inequality is the concept most commonly used in the literature when dealing with the analysis of inequality and is the focus of this study.

5.2.2.2 Measurement

Desirable features of an inequality measure

It is required that an inequality measure should meet the following five key axioms (Litchfield 1999 and Haughton and Khandker 2009: 105-106):

- *Pigou-Dalton transfer principle*: An income transfer from a poorer person to a richer person should result in an increase (or least not a decrease) in inequality. Similarly, a decline (or at least not a rise) in inequality takes place if there is an income transfer from a richer person to a poorer person.
- *Income scale independence*: The inequality should not be influenced by the magnitude of total income. That is, the inequality measure should not change if everyone's income increases or decreases by the same proportion.
- *Principle of population*: The inequality measure should not depend on the number of income receivers. For example, assuming there are three people receiving income in country A and each of them earn R1 000, while in country B there are five people receiving income with each of them earning R500, the inequality measures should be

[93] The following example illustrates the difference between the two types of inequality: Assuming there are two households in the economy, one with an income of R1 000 and the other with an income of R10 000. If the income of both households increases at the same rate of 100% during a period, the new household income would be R2 000 and R20 000 respectively. Absolute inequality increases as the difference in their incomes has risen from R9 000 to R18 000. In contrast, relative inequality remains unchanged as the richer household is still 10 times richer (i.e., R10 000 / R1 000 = R20 000 / R2 000 = 10).

the same in both countries (perfect equality in this example).

- *Anonymity*: Inequality should only be affected by the incomes of the individuals, not by their other characteristics.
- *Decomposability*: This requires overall inequality to be related consistently to constituent parts of the distribution, such as population sub-groups (e.g., race, gender). For example, assuming inequality rises amongst each sub-group of the population, it is expected that overall inequality would also increase.

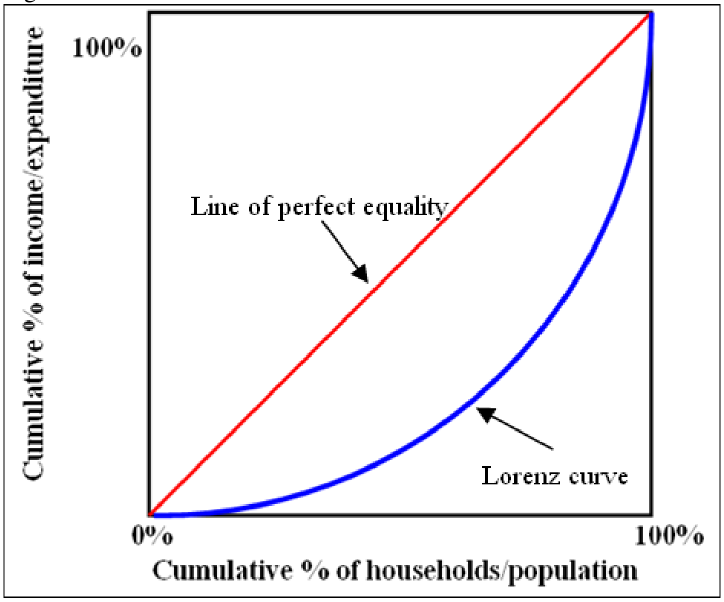
One of the ways to categorize the measures of inequality is to distinguish indecisive from decisive measures (Whiteford and McGrath 1994: 6). Indecisive measures do not attempt to summarize the income / consumption / expenditure distribution into a single coefficient, for example, the share of income accruing to each population group or gender group, the shares of total income accruing to specified percentiles of the population (e.g., the first quintile or lowest 20%, and the tenth decile or highest 10%). In contrast, the decisive measures provide information about the distribution of income in the form of a single coefficient, for example, the Gini coefficient, the Theil entropy index and Atkinson's measure. These three measures will be discussed in greater detail below.

Inequality measure: Gini coefficient

The Gini coefficient is probably the most commonly used measure of inequality and can be derived with the aid of a Lorenz curve. To construct the Lorenz curve illustrating the income distribution, the population first have to be ranked from poorest to richest and this is done on a cumulative percentage basis. In other words, the poorest percent of the population is started with, followed by the second poorest percent and so on. The cumulative percentages of the population are then plotted on the horizontal axis. The vertical axis shows the cumulative percentage of total income.

The Gini coefficient is equivalent to the area between the diagonal and the Lorenz curve (actual deviation from equality) divided by the area under the diagonal (maximum possible deviation from equality), as shown in Figure 5.5. The Gini coefficient is ranged between zero and one: if everyone earns exactly the same income, then the Lorenz curve would coincide with the line of perfect equality, and the Gini coefficient would be zero. In contrast, if one person earns all the income, the coefficient would be (almost) equal to one.

Figure 5.5: The Lorenz curve



Although an advantage of the Gini coefficient is that it is easy to interpret, a major drawback is that it is not decomposable, i.e., the Gini coefficient only meets the first four axioms as discussed above (Litchfield 1999). Hence, an alternative way to measure inequality, namely the Theil's generalized entropy (GE) measures, is considered.

Inequality measure: GE measures

The general formula of the GE measures is given by: $GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{\alpha} - 1\right]$,

where N stands for the population size, y_i is the per capita income of the i-th person, and $\bar{y}$ is the mean per capita income. The parameter α represents weights given to distances between incomes at different parts of the income distribution. GE is more sensitive to changes in the lower tail of the distribution for lower values of α. In contrast, for higher values of α, GE is more sensitive to changes that affect the upper tail (Haughton and Khandker, 2009: 106-107). The most commonly used values of α are 0 and 1, with GE(0) and GE(1) being known as Theil-L and Theil-T indices respectively. GE(0) is also known as mean log deviation. The respective formulae of the two indices are as follows:

- Theil-L index $= GE(0) = \frac{1}{N}\sum_{i=1}^{N}\frac{\bar{y}}{y_i}$
- Theil-T index $= GE(1) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\bar{y}}\ln\left(\frac{y_i}{\bar{y}}\right)$

An advantage of the Theil measures over the Gini coefficient is that it is decomposable – the total inequality of a population can be broken down into a weighted average of the inequality existing within subgroups of the population (e.g., race) and the inequality existing between the subgroups. The breakdown of each measure is as follows:

- $GE(0) = \sum_j \left(\frac{N_j}{N}\right) GE(0)_j + \sum_j \left(\frac{N_j}{N}\right) \ln\left(\frac{\overline{y}}{\overline{y_j}}\right)$, where N once again stands for population size, N_j is the population size of a subgroup, $GE(0)_j$ is the Theil-L index of this subgroup, $\overline{y}$ is the mean per capita income of the population, and $\overline{y_j}$ is the mean per capita income of the subgroup concerned. The first term of the equation represents within-group inequality, while the second term measures between-group inequality.
- $GE(1) = \sum_j \left(\frac{Y_j}{Y}\right) GE(1)_j + \sum_j \left(\frac{Y_j}{Y}\right) \ln\left(\frac{Y_j / Y}{N_j / N}\right)$, where Y_j and Y stand for the total income of a subgroup and the population respectively, and $GE(1)_j$ represents the Theil-T index of the subgroup. The first and second components of the above equation measures within-group inequality and between-group inequality respectively.

Although the Theil indices are decomposable (in fact, the Theil measures meet all five axioms as discussed above), the indices are more difficult to interpret compared with the Gini coefficient, since the former indices do not have an upper limit.

<u>Inequality measure: Atkinson's measure</u>

Atkinson (1970) proposed another inequality measure that also includes a weight parameter (ε) as in the two Theil measures. The Atkinson's inequality index is given as:

$$A_\varepsilon = 1 - \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\overline{y}}\right)^{1-\varepsilon}\right]^{1/(1-\varepsilon)} \quad \text{, with } 0 \le \varepsilon < \infty$$

The parameter ε measures aversion to inequality. The higher the value of ε, the more concerned the society is about inequality. That is, if society is indifferent about inequality, ε would be set to zero; when ε is set to infinity, it means the society is extremely concerned with the poorest households. The Atkinson's measure, which is ranged from zero and one, meets all axioms except the decomposability requirement[94].

Growth incidence curve

A growth incidence curve (GIC) provides the rate of growth in per capita income over a certain time period at each percentile of the distribution, ranked by per capita income or expenditure (Ravallion 2004). The following procedures are involved when the GIC is derived (Haughton and Khandker 2009):

- Dividing the data from the first survey (e.g., IES 1995) into percentiles, and computing income per capita for each of the percentiles.
- Dividing the data from the second survey (e.g., IES 2000) into percentiles, and again computing income per capita for each of the percentiles.
- After adjusting for inflation across the two periods, compute the percentage change in real per capita income for each percentile and graph the results.

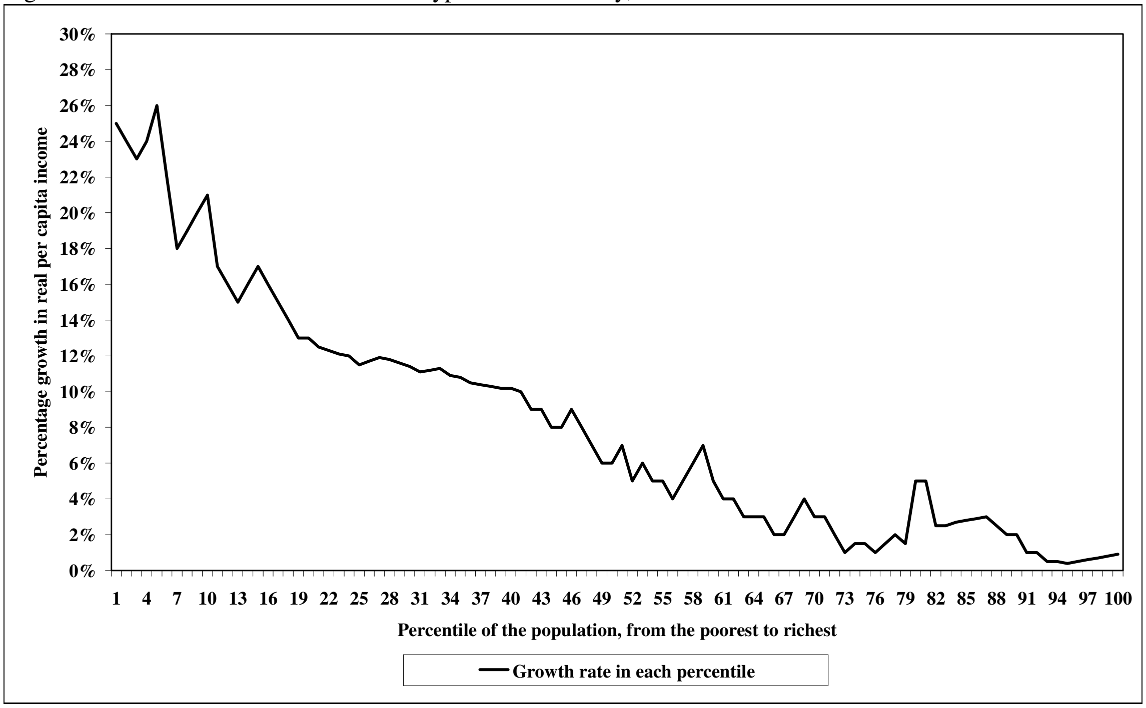
Bhorat and Van der Westhuizen (2008) argue that pro-poor growth may be defined in two ways:

- Growth is pro-poor in an absolute sense if the change in income levels of the poor, as defined by a chosen poverty line, over a period of time is larger than zero. That is, the income levels of the poor have increased in absolute terms. This is represented graphically by a GIC that is located above zero along the whole distribution.
- Growth is pro-poor in a relative sense if the change in the income levels of the poor is larger than that of the non-poor. This is represented graphically by a GIC that is downward-sloping.

Figure 5.6 illustrates a GIC of a hypothetical country between 1995 and 1996 that reflects pro-poor growth in both absolute sense and relative sense, as indicated by the fact that the GIC lies above the horizontal axis and is downward-sloping in general. This pro-poor growth would consequently result in reduction of both income poverty and inequality.

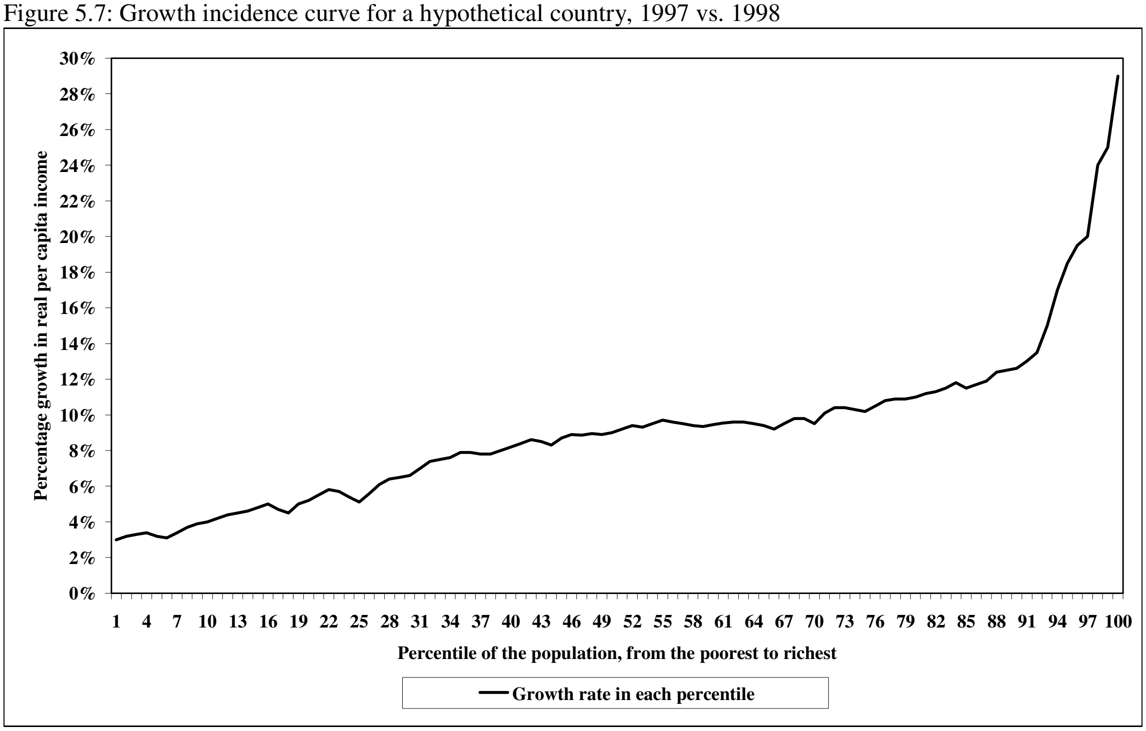
[94] Nevertheless, the measure can be pseudo-decomposed into within- and between-group inequality, but the sum of these two components does not sum exactly to the overall measure, leaving a residual (Woolard, 2001: 153-154).

Figure 5.6: Growth incidence curve for a hypothetical country, 1995 vs. 1996



In contrast, Figure 5.7 shows a GIC that reflects pro-poor growth only in absolute sense. Although the income at all percentiles of a hypothetical country increased between 1997 and 1998, the growth was relatively slower for those at the bottom of the income distribution. This income growth would still lead to poverty reduction, but income inequality would worsen.

Figure 5.7: Growth incidence curve for a hypothetical country, 1997 vs. 1998



GICs will be applied across the household surveys, when poverty trends are looked at in Chapter 6.

5.3 Applying sequential regression multiple imputation (SRMI) on surveys with zero or unspecified income or expenditure

As discussed in Section 3.8, a serious problem in some surveys is the presence of a high proportion of households reporting zero or unspecified income. If the zero-income households were included, this would result in a higher poverty and inequality estimates. In contrast, excluding the households with unspecified income from the analysis would result in an over-estimation or under-estimation of poverty and inequality, depending on whether these households come from the bottom or top of the income distribution. Hence, the SRMI method is applied to deal with households with zero or unspecified reported income (or expenditure) category in the censuses, CS 2007, GHSs, and OHSs/LFSs. This section aims to discuss in greater detail how the method is applied in each survey.

First, Table 5.8 explains the decision rules on the personal income variable, before SRMI1 (i.e., imputations are person level) was applied. In Census 2001, the personal income variable without hot deck imputation by Stats SA was used. The employment status of the person was the critical variable that was taken into consideration before deciding whether to accept his/her declared personal income category, or whether to adjust his/her personal income to missing, before SRMI1 was run. To summarize, for people aged 15 years or above who were employed at the time of the survey but declared zero or unspecified personal income, their personal income was assumed to be unspecified, and SRMI1 was applied to impute their personal income category.

Only the employed were included for the SRMI1, and the SRMI1 was run five times. The variables included for the SRMI1 were as follows: Race, gender, province, age, years of educational attainment, broad occupation category of employed, broad industry category of employed, number of employed in the household, and annual personal income category. Person weight was the weight variable. The average of the five imputed personal income values was regarded as the final imputed personal income. Finally, the household income amount was derived as the sum of the personal income amounts, i.e., households falling under the same household income category could have different household income amounts.

Table 5.8: Decision rules before SRMI1 is applied on the personal income variable

<u>Census 1996</u>
○ If additional household income or remittances received were unspecified, they were set to R0.
○ If personal income was specified as R0:
– If age was 0-14 years, then personal income remained R0, regardless of labour status
– If age was 15+ years and the labour status was given as unemployed/inactive/not working-age population, the personal income remained R0
– If age was 15+ years and the labour status was given as employed, then personal income became unspecified, and was imputed by SRMI1
– If age was unspecified, then personal income remained R0, regardless of labour status
○ If personal income was unspecified:
– If age was 0-14 years, then personal income became R0, regardless of labour status
– If age was 15+ years and the labour status was given as unemployed/inactive/not working-age population, the personal income became R0
– If age was 15+ years and the labour status was given as employed, then personal income remained unspecified, and was imputed by SRMI1
– If age was unspecified, then personal income became R0, regardless of labour status
○ After that, the personal incomes + additional household income + remittances received were added to derive the household income amount, before per capita income was derived
<u>Census 2001</u>
○ Personal income without hot deck imputation was used
○ If personal income before hot deck imputation was specified as R0:
– If age was 0-14 years, then personal income remained R0, regardless of labour status
– If age was 15+ years and the labour status was given as unemployed/inactive/not working-age population, the personal income remained R0
– If age was 15+ years and the labour status was given as employed, then personal income became unspecified, and was imputed by SRMI1
– If age was unspecified, then personal income remained R0, regardless of labour status
○ If personal income before hot deck imputation was unspecified:
– If age was 0-14 years, then personal income became R0, regardless of labour status
– If age was 15+ years and the labour status was given as unemployed/inactive/not working-age population, the personal income became R0
– If age was 15+ years and the labour status was given as employed, then personal income remained unspecified, and was imputed by SRMI1
– If age was unspecified, then personal income became R0, regardless of labour status
○ After that, the personal incomes were added to derive the household income amount, before per capita income was derived
<u>CS 2007</u>
○ If personal income was specified as R0:
– If age was 0-14 years, then personal income remained R0, regardless of labour status
– If age was 15+ years and the labour status was given as unemployed/inactive/not working-age population, the personal income remained R0
– If age was 15+ years and the labour status was given as employed, then personal income became unspecified, and was imputed by SRMI1
– If age was unspecified, then personal income remained R0, regardless of labour status
○ If personal income was unspecified:
– If age was 0-14 years, then personal income became R0, regardless of labour status
– If age was 15+ years and the labour status was given as unemployed/inactive/not working-age population, the personal income became R0
– If age was 15+ years and the labour status was given as employed, then personal income remained unspecified, and was imputed by SRMI1
– If age was unspecified, then personal income became R0, regardless of labour status
○ After that, the personal incomes were added to derive the household income amount, before per capita income was derived

Table 5.9: Number of people and households in each monthly personal income and household income category respectively, before and after SRMI1 was applied

	Before SRMI1		**After SRMI1**			
	Personal income		**Personal income**		**Household income**	
Census 1996						
1: None	22 638 513	60.6%	26 022 127	69.7%	1 284 285	14.8%
2: R1 – R200	1 013 994	2.7%	1 014 712	2.7%	600 928	6.9%
3: R201 – R500	3 127 647	8.4%	3 153 716	8.4%	1 507 158	17.3%
4: R501 – R1 000	1 778 993	4.8%	1 841 289	4.9%	1 189 838	13.7%
5: R1001 – R1 500	1 461 100	3.9%	1 531 381	4.1%	972 733	11.2%
6: R1 501 – R2 500	1 255 632	3.4%	1 321 943	3.5%	899 576	10.3%
7: R2 501 – R3 500	749 239	2.0%	810 986	2.2%	512 606	5.9%
8: R3 501 – R4 500	494 498	1.3%	548 146	1.5%	393 412	4.5%
9: R4 501 – R6 000	458 961	1.2%	495 662	1.3%	419 307	4.8%
10: R6 001 – R8 000	237 232	0.6%	256 541	0.7%	288 145	3.3%
11: R8 001 – R11 000	159 170	0.4%	166 930	0.4%	294 660	3.4%
12: R11 001 – R16 000	96 327	0.3%	98 146	0.3%	184 037	2.1%
13: R16 001 – R30 000	57 637	0.2%	57 862	0.2%	123 657	1.4%
14: R30 001 or more	22 032	0.1%	22 042	0.1%	36 137	0.4%
99: Unspecified	3 790 508	10.2%	0	0.0%	0	0.0%
	37 341 483	100.0%	37 341 483	100.0%	8 706 479	100.0%
Census 2001						
1: None	23 434 110	56.1%	29 247 806	70.1%	2 673 559	24.7%
2: R1 – R400	2 046 913	4.9%	2 053 857	4.9%	860 093	7.9%
3: R401 – R800	3 663 976	8.8%	3 778 178	9.1%	1 894 392	17.5%
4: R801 – R1 600	2 008 797	4.8%	2 182 107	5.2%	1 689 132	15.6%
5: R1 601 – R3 200	1 706 388	4.1%	1 873 328	4.5%	1 386 097	12.8%
6: R3 201 – R6 400	1 263 542	3.0%	1 398 279	3.3%	988 268	9.1%
7: R6 401 – R12 800	677 332	1.6%	762 120	1.8%	706 331	6.5%
8: R12 801 – R25 600	256 999	0.6%	285 743	0.7%	412 061	3.8%
9: R25 601 – R51 200	89 543	0.2%	94 449	0.2%	144 288	1.3%
10: R51 201 – R102 400	35 182	0.1%	35 611	0.1%	37 414	0.3%
11: R102 401 – R204 800	25 877	0.1%	25 877	0.1%	23 278	0.2%
12: R204 801 or more	9 859	0.0%	9 859	0.0%	13 576	0.1%
13: Unspecified	6 528 696	15.6%	0	0.0%	0	0.0%
	41 747 214	100.0%	41 747 214	100.0%	10 828 489	100.0%
CS 2007						
1: None	22 058 265	46.6%	22 926 594	48.4%	1 069 905	8.6%
2: R1 – R400	7 967 281	16.8%	7 970 421	16.8%	610 223	4.9%
3: R401 – R800	2 342 025	4.9%	2 494 369	5.3%	1 105 489	8.9%
4: R801 – R1 600	5 660 829	11.9%	6 132 539	12.9%	2 431 775	19.6%
5: R1 601 – R3 200	2 274 924	4.8%	2 780 130	5.9%	2 628 573	21.2%
6: R3 201 – R6 400	1 808 507	3.8%	2 154 224	4.5%	1 772 450	14.3%
7: R6 401 – R12 800	1 413 691	3.0%	1 647 038	3.5%	1 214 057	9.8%
8: R12 801 – R25 600	654 204	1.4%	778 886	1.6%	847 908	6.8%
9: R25 601 – R51 200	283 171	0.6%	321 326	0.7%	463 795	3.7%
10: R51 201 – R102 400	88 590	0.2%	96 085	0.2%	152 809	1.2%
11: R102 401 – R204 800	46 329	0.1%	46 470	0.1%	48 020	0.4%
12: R204 801 or more	26 519	0.1%	26 519	0.1%	33 752	0.3%
13: Unspecified	2 750 266	5.8%	0	0.0%	0	0.0%
	47 374 601	100.0%	47 374 601	100.0%	12 378 756	100.0%

The second and third columns of Table 5.9 above show the percentage of people in each personal income category in each survey, when the discussed decision rules were adopted before the SRMI1 was run. In addition, in all three surveys, more than 90% of people with zero reported personal income were not employed at the time of the survey. However, with regard to people with unspecified personal income, approximately 15% of such people in each census were employed, but this proportion was very high in CS 2007 (58.5%), and a higher proportion of them were whites (22.7%, compared with about 15% in the two censuses). This implies that a lot of employed in CS 2007 refused to specify their personal income and excluding them would have resulted in over-estimation of poverty. The last four columns of Table 5.9 show the percentage of people/households in each personal/household income category in each survey after SRMI1. The percentage of households with zero household income in each survey is 14.8%, 24.7% and 8.6% in 1996, 2001 and 2007 respectively, after SRMI1 was run.

As mentioned earlier, the labour market status of the person (i.e., whether he/she was employed or not at the time of the survey) is an important factor in determining whether to accept the person's declared personal income. However, when looking at Figures A.1-A.4 in Appendix A, it can be seen that the labour market status of the working-age population was not captured particularly well in the three surveys, especially in 2001 and 2007 (i.e., under-estimation of labour force participation rate but large over-estimation of unemployment rate), compared with LFSs taking place during the same year.

The main aim of the two censuses as well as CS 2007 was not to capture labour market status of the respondents. In fact, only very few questions (approximately five in each survey) were asked on the labour market activities of the respondents. Therefore, running SRMI on unspecified personal income of the employed might not be the best approach. Therefore, the SRMI was run at household level (i.e., SRMI2).

As discussed in Section 2.2, household income was derived differently across the three surveys, and the household income variable derived by Stats SA in Census 1996 is incorrect as discussed earlier (refer to Table 2.4). In addition, it could be argued that, after applying the SRMI at person level, there still remains a high proportion of households with zero household income (14.8% in 1996, 24.7% in 2001, and 8.6% in 2007, as seen in Table 5.9). Most of these households should have some sources of non-work-related income (i.e., remittances from other members, social grants, etc.), or they would not be able to survive. Therefore, in this section, the SRMI is run at household level, i.e., SRMI2. However, before running SRMI2, a consistent method must be applied to derive household income across the three surveys. This method is presented in Table 5.10.

Table 5.10: A consistent approach to derive the household income variable before SRMI2 was applied

Census 1996

- If additional household income or remittances received were unspecified, they were set to R0.
- Household income was derived by adding personal income of all members, additional household income, and remittances received.
- If personal income of any member was unspecified:
  - If age was 0-14 years, then personal income was set to zero
  - If age was unspecified, then personal income was set to zero
  - If age was 15+ years, personal income remained unspecified, and households containing at least 1 such person would have unspecified household income. SRMI2 was needed later.

Census 2001

- Personal income without hot deck imputation was used
- Household income was derived by adding the personal income of all members
- If personal income of any member was unspecified:
  - If age was 0-14 years, then personal income was set to zero
  - If age was unspecified, then personal income was set to zero
  - If age was 15+ years, personal income remained unspecified, and households containing at least 1 such person would have unspecified household income. SRMI2 was needed later.

CS 2007

- Household income was derived adding the personal income of members.
- If personal income of any member was unspecified:
  - If age was 0-14 years, then personal income was set to zero
  - If age was unspecified, then personal income was set to zero
  - If age was 15+ years, personal income remained unspecified, and households containing at least 1 such person would have unspecified household income. SRMI2 was needed later.

Once this consistent pre-SRMI2 household income was derived, a decision had to be made on how to deal with households with zero or unspecified household income. It was decided to apply further decision rules (as they will be referred to for the remainder of the dissertation), before SRMI2 was eventually run (See Table 5.11).

Table 5.11: Further decision rules on the derivation of household income before SRMI2 was applied

Census 1996, Census 2001 and CS 2007

- If the household income was non-zero, accept it.
- If the household income was R0, but it did not contain any member 15+ years with R0 personal income, this R0 household income was accepted.
- If the household income was R0, and it contained at least 1 person 15+ years with R0 personal income, then the household income became unspecified. SRMI2 was applied.
- If the household income was unspecified, SRMI2 was applied.

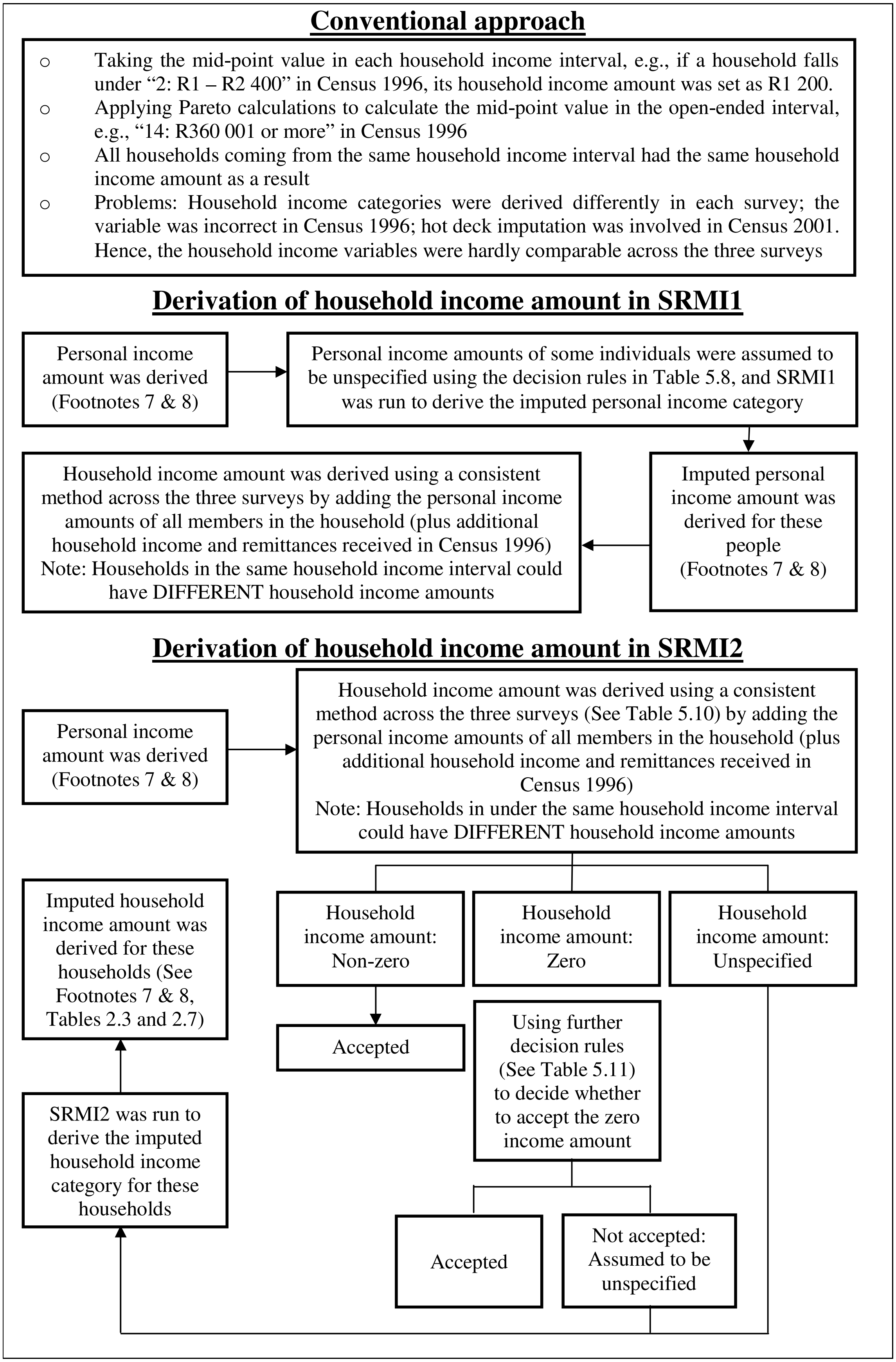
Thus, it can be seen that, unlikely SRMI1, the labour market status of the person is no longer an important concern when running SRMI2. In addition, the variables included for the SRMI2 (household weight was the weight variable) were as follows: Province, race of household head, gender of household head, age of household head, years of educational attainment of household head, employment status of household head, number of employed in the household, household size, and annual household income category.

Table 5.12 presents the percentage of households in each annual household income category in each survey, before and after the application of SRMI2. In all three surveys, fewer than 1% of households had zero estimated household income in all three surveys after the imputations.

Table 5.12: Number of households in each monthly household income category, before and after SRMI2

	Before SRMI2 + Before further decision rules were applied		Before SRMI2 + After further decision rules were applied		After SRMI2	
Census 1996						
1: None	1 106 024	12.8%	58 186	0.7%	58 186	0.7%
2: R1 – R200	553 564	6.4%	553 564	6.4%	590 476	6.9%
3: R201 – R500	1 392 516	16.2%	1 392 516	16.2%	1 935 377	22.5%
4: R501 – R1 000	1 066 666	12.4%	1 066 666	12.4%	1 667 011	19.3%
5: R1001 – R1 500	841 956	9.8%	841 956	9.8%	1 144 259	13.3%
6: R1 501 – R2 500	771 575	9.0%	771 575	9.0%	943 928	11.0%
7: R2 501 – R3 500	431 527	5.0%	431 527	5.0%	544 755	6.3%
8: R3 501 – R4 500	329 892	3.8%	329 892	3.8%	414 470	4.8%
9: R4 501 – R6 000	350 640	4.1%	350 640	4.1%	416 054	4.8%
10: R6 001 – R8 000	238 586	2.8%	238 586	2.8%	289 670	3.4%
11: R8 001 – R11 000	246 922	2.9%	246 922	2.9%	281 943	3.3%
12: R11 001 – R16 000	155 518	1.8%	155 518	1.8%	175 937	2.0%
13: R16 001 – R30 000	107 971	1.3%	107 971	1.3%	118 628	1.4%
14: R30 001 or more	32 259	0.4%	32 259	0.4%	36 645	0.4%
99: Unspecified	991 723	11.5%	2 039 561	23.7%	0	0.0%
	8 617 339	100.0%	8 617 339	100.0%	8 617 339	100.0%
Census 2001						
1: None	2 274 882	21.0%	13 567	0.1%	13 567	0.1%
2: R1 – R400	774 583	7.2%	774 583	7.2%	838 221	7.7%
3: R401 – R800	1 686 640	15.6%	1 686 640	15.6%	3 212 187	29.7%
4: R801 – R1 600	1 437 798	13.3%	1 437 798	13.3%	2 751 117	25.4%
5: R1 601 – R3 200	1 119 402	10.3%	1 119 402	10.3%	1 604 993	14.8%
6: R3 201 – R6 400	759 920	7.0%	759 920	7.0%	1 038 319	9.6%
7: R6 401 – R12 800	529 351	4.9%	529 351	4.9%	728 931	6.7%
8: R12 801 – R25 600	302 734	2.8%	302 734	2.8%	415 782	3.8%
9: R25 601 – R51 200	107 869	1.0%	107 869	1.0%	148 392	1.4%
10: R51 201 – R102 400	29 814	0.3%	29 814	0.3%	41 728	0.4%
11: R102 401 – R204 800	19 051	0.2%	19 051	0.2%	22 826	0.2%
12: R204 801 or more	11 038	0.1%	11 038	0.1%	12 426	0.1%
13: Unspecified	1 775 407	16.4%	4 036 722	37.3%	0	0.0%
	10 828 489	100.0%	10 828 489	100.0%	10 828 489	100.0%
CS 2007						
1: None	1 022 550	8.3%	5 940	0.0%	5 940	0.0%
2: R1 – R400	623 073	5.0%	623 073	5.0%	625 313	5.1%
3: R401 – R800	1 118 947	9.0%	1 118 947	9.0%	1 329 209	10.7%
4: R801 – R1 600	2 366 175	19.1%	2 366 175	19.1%	2 998 839	24.2%
5: R1 601 – R3 200	2 391 387	19.3%	2 391 387	19.3%	2 892 378	23.4%
6: R3 201 – R6 400	1 438 767	11.6%	1 438 767	11.6%	1 786 791	14.4%
7: R6 401 – R12 800	965 259	7.8%	965 259	7.8%	1 211 580	9.8%
8: R12 801 – R25 600	677 998	5.5%	677 998	5.5%	851 889	6.9%
9: R25 601 – R51 200	363 323	2.9%	363 323	2.9%	448 032	3.6%
10: R51 201 – R102 400	121 188	1.0%	121 188	1.0%	148 828	1.2%
11: R102 401 – R204 800	42 297	0.3%	42 297	0.3%	48 623	0.4%
12: R204 801 or more	29 821	0.2%	29 821	0.2%	31 334	0.3%
13: Unspecified	1 217 971	9.8%	2 234 581	18.1%	0	0.0%
	12 378 756	100.0%	12 378 756	100.0%	12 378 756	100.0%

Figure 5.8: A summary of the derivation of household income amount under various methods in the two censuses and CS 2007



Finally, as far as the derivation of household income is concerned, for households that did not require SRMI2, the household income amount equals the sum of the personal income amounts, and households from the same household income category could have different household income amounts. However, with regard to the households with imputed household income category after SRMI2, they had their household income amount derived using the values in Tables 2.1 and 2.5. For example, if the imputed household income category of a household in Census 1996 is "1: R1 – R2 400" after SRMI2, then the annual household income amount is approximated as R1 600. Similarly, if the imputed household income category of a household in Census is "5: R19 201 – R38 400", then the annual household income amount is estimated as R27 153.

Figure 5.8 above summarizes the approaches to derive the household income amount in the two censuses and CS 2007.

As discussed in Sections 2.4.2.3 and 2.5.2.3, the OHSs/LFSs and GHSs also contained households with unspecified income or expenditure. Hence, SRMI at household level (i.e., SRMI2) was also applied to impute the household expenditure amount (in OHS 1996-1998) and household income or expenditure category (OHS 1999 and LFS 2001-2004) respectively, and the results are presented in Tables 5.13 and 5.14.

Table 5.13: Proportion of households in each monthly household income or expenditure category after SRMI2, OHSs/LFSs

	OHS					**LFS (September)**			
	1996	**1997**	**1998**	**1999**	**1999 (Income)**	**2001**	**2002**	**2003**	**2004**
R0 – R399	21.5%	24.1%	26.6%	25.1%	15.9%	31.8%	30.6%	24.6%	21.8%
R400 – R799	26.4%	30.7%	30.2%	27.9%	22.0%	26.9%	26.7%	29.0%	28.8%
R800 – R1199	15.8%	15.5%	13.2%	15.0%	13.6%	12.2%	12.6%	14.3%	14.4%
R1200 – R1799	11.0%	9.3%	8.4%	9.6%	12.8%	7.7%	7.7%	7.8%	8.8%
R1800 – R2499	8.1%	5.8%	6.5%	6.9%	8.3%	5.8%	6.2%	6.1%	6.9%
R2500 – R4999	12.1%	9.5%	10.0%	8.9%	12.3%	8.7%	8.0%	8.9%	8.7%
R5000 – R9999	4.2%	4.1%	4.1%	5.0%	8.7%	5.1%	6.0%	6.2%	7.2%
R10000 or more	1.0%	1.1%	0.9%	1.7%	6.3%	1.8%	2.3%	3.1%	3.3%
Don't know / Refuse / Unspecified	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
R0 – R1199	63.7%	70.3%	70.0%	68.0%	51.5%	70.4%	69.9%	67.9%	65.0%

Table 5.14: Proportion of households in each monthly household expenditure category after SRMI2, GHSs

	2002	2003	2004	2005	2006	2007	2008	2009
R0	N/A							0.5%
R1 – R199								2.4%
R200 – R399								8.4%
R0 – R399	31.4%	26.0%	18.5%	19.2%	17.6%	13.5%	9.4%	11.3%
R400 – R799	27.8%	28.2%	29.0%	28.3%	29.1%	27.4%	23.3%	20.1%
R800 – R1199	12.6%	14.2%	14.6%	15.4%	17.7%	18.2%	19.7%	17.8%
R1200 – R1799	7.6%	8.3%	10.7%	10.6%	10.7%	12.2%	12.8%	13.3%
R1800 – R2499	5.9%	6.3%	7.1%	6.7%	6.7%	7.3%	8.9%	9.6%
R2500 – R4999	7.7%	8.5%	10.5%	10.6%	9.4%	10.9%	11.8%	11.2%
R5000 – R9999	5.3%	6.0%	7.2%	6.6%	6.2%	7.5%	8.5%	9.7%
R10000 or more	1.8%	2.5%	2.4%	2.7%	2.7%	3.1%	5.6%	7.1%
Don't know / Refuse / Unspecified	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***	***0.0%***
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
R0 – R1199	71.8%	68.4%	62.1%	62.9%	64.4%	59.1%	52.4%	49.1%

5.4 Derivation of the real per capita income, expenditure and consumption variables

This section discusses the derivation of the per capita income, expenditure and consumption variables to be used for the forthcoming poverty and inequality analyses. These variables were simply derived as the total household income or expenditure or consumption divided by the household size, if the former variables were reported as actual amounts. In OHS 1999, LFS 2001-2004, GHS 2002-2008, AMPS 1993-2009 and OHS 1999 (household income or expenditure being a categorical variable), the total amount was approximated as the mid-point of the class interval of each category. On the other hand, to determine the midpoint value of the open category (R10 000 or more), a Pareto function was fitted. That is, the midpoint-Pareto method (1) as discussed in Section 3.6.2 was applied. The total income or expenditure amount was then divided by the household size to derive the per capita income or expenditure amount.

Finally, all nominal amounts were converted into real per capita income 2000 prices, using the South African Reserve Bank's monthly CPI series (Data code in the Quarterly Bulletin of Reserve Bank: 7032). The CPI values used in each survey are shown in Table 5.15.

Table 5.15: Monthly CPIs (Data code in the Quarterly Bulletin of Reserve Bank: 7032) used to convert the nominal values into 2000 prices

Survey	Year	Month	CPI	Survey	Year	Month	CPI
Census	1996	Oct	79.78	NIDS	2008	Jan	151.20
Census	2001	Oct	106.05	NIDS	2008	Feb	152.06
Census	2007	Feb	138.45	NIDS	2008	Mar	153.86
IES	1995	Oct	73.20	NIDS	2008	Apr	156.01
IES	2000	Oct	101.96	NIDS	2008	May	158.36
IES	2005	Sep	129.25	NIDS	2008	Jun	160.32
IES	2005	Oct	129.64	NIDS	2008	Jul	163.10
IES	2005	Nov	129.88	NIDS	2008	Aug	164.30
IES	2005	Dec	130.32	NIDS	2008	Sep	164.80
IES	2006	Jan	130.56	NIDS	2008	Oct#	165.00
IES	2006	Feb	130.91	NIDS	2008	Nov#	165.80
IES	2006	Mar	131.18	NIDS	2008	Dec#	165.80
IES	2006	Apr	131.40	NIDS	2008	Missing##	160.05
IES	2006	May	132.30	PSLSD	1993	Jul	61.73
IES	2006	Jun	133.52	AMPS###	1993	Jan-Dec	61.16
IES	2006	Jul	134.44	AMPS	1994	Jan-Dec	66.63
IES	2006	Aug	135.70	AMPS	1995	Jan-Dec	72.41
OHS	1996	Nov	80.42	AMPS	1996	Jan-Dec	77.73
OHS	1997	Oct	85.87	AMPS	1997	Jan-Dec	84.42
OHS	1998	Oct	93.57	AMPS	1998	Jan-Dec	90.22
OHS	1999	Oct	95.27	AMPS	1999	Jan-Dec	94.90
LFS	2001	Sep	106.00	AMPS	2000	Jan-Dec	99.97
LFS	2002	Sep	117.94	AMPS	2001	Jul-Dec	106.49
LFS	2003	Sep	122.18	AMPS	2002	Jul-Dec	118.86
LFS	2004	Sep	123.86	AMPS	2003	Jul-Dec	122.04
GHS	2002	Jul	115.89	AMPS	2004	Jan-Dec	123.80
GHS	2003	Jul	121.96	AMPS	2005	Jan-Jun	126.61
GHS	2004	Jul	123.89	AMPS	2006	Jan-Jun	131.56
GHS	2005	Jul	128.06	AMPS	2007	Jul-Dec	146.77
GHS	2006	Jul	134.44	AMPS	2008	Jan-Jun	155.32
GHS	2007	Jul	143.83	AMPS	2009	Jan-Dec	168.84
GHS	2008	Jul	163.10				
GHS	2009	Jul#	173.40				

New weights (using the results from IES 2005-2006) were adopted for the derivation of the CPI since November 2008 (Data code in the Quarterly Bulletin of Reserve Bank: 7170), and using the new weights, the CPI was derived only from 2002. The monthly CPI values in the last three months of 2008 amounted to 102.6, 103.1 and 103.1 respectively, and thus the inflation between October and November is 0.49% and the inflation between November and December is 0%. Hence, the CPI in November and December 2008 using the old weights (i.e., code: 7032) could be approximated as 165.80 [165.00 × (1 + 0.49%)] and 165.80 respectively [165.00 × (1 + 0%)]. A similar approach was adopted when estimating the CPI for GHS 2009 using the old weights.

7 households (out of 7 305) in NIDS had missing interview month, and the average of the 12 monthly CPIs in 2008 (i.e., 160.06) was used to convert the nominal amounts into 2000 prices in these households.

As the AMPS metadata did not specify the exact survey month, if the AMPS took place twice a year (i.e., 1993 – 2000, 2004 and 2009), the annual CPI was used to deflate the nominal per capita income. However, if only one AMPS took place in the first half of the year (i.e., 2005, 2006 and 2008), the average of the January-June monthly CPI values was used to deflate the nominal per capita income. Similarly, the average of the July-December CPI values was used if only one AMPS took place in the second half of the year (i.e., 2001 – 2003 and 2007).

Figures 5.9 and 5.10 show the mean per capita income/expenditure/consumption (2000 prices, per annum) in each survey, and as expected, the mean values clearly increased after imputations were applied on the censuses, CS 2007, LFS 2003b and LFS 2004b, but showed negligible change in all GHSs and OHSs, as well as LFS 2001b and LFS 2002b. Furthermore, the AMPS per capita values are higher than the OHS/LFS/GHS values of the same year.

Figure 5.9: Mean annual per capita income/expenditure/consumption (2000 prices) in the IESs, PSLSD, OHS 1999 (Income), NIDS, censuses and CS 2007

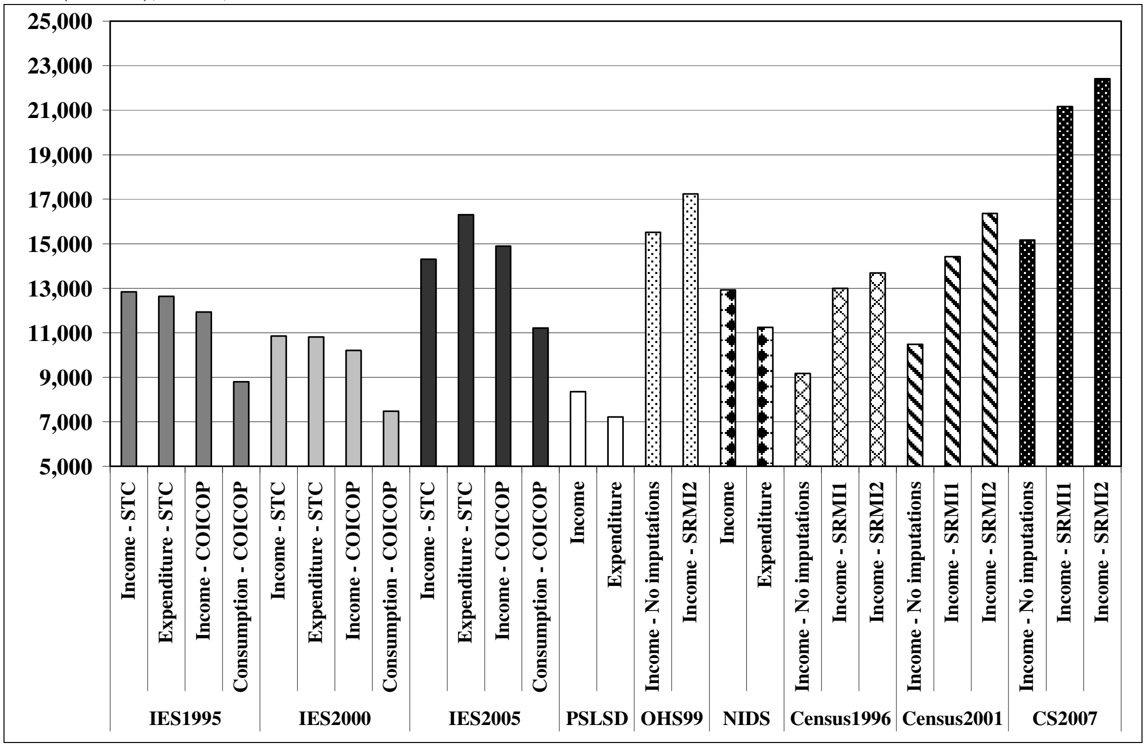
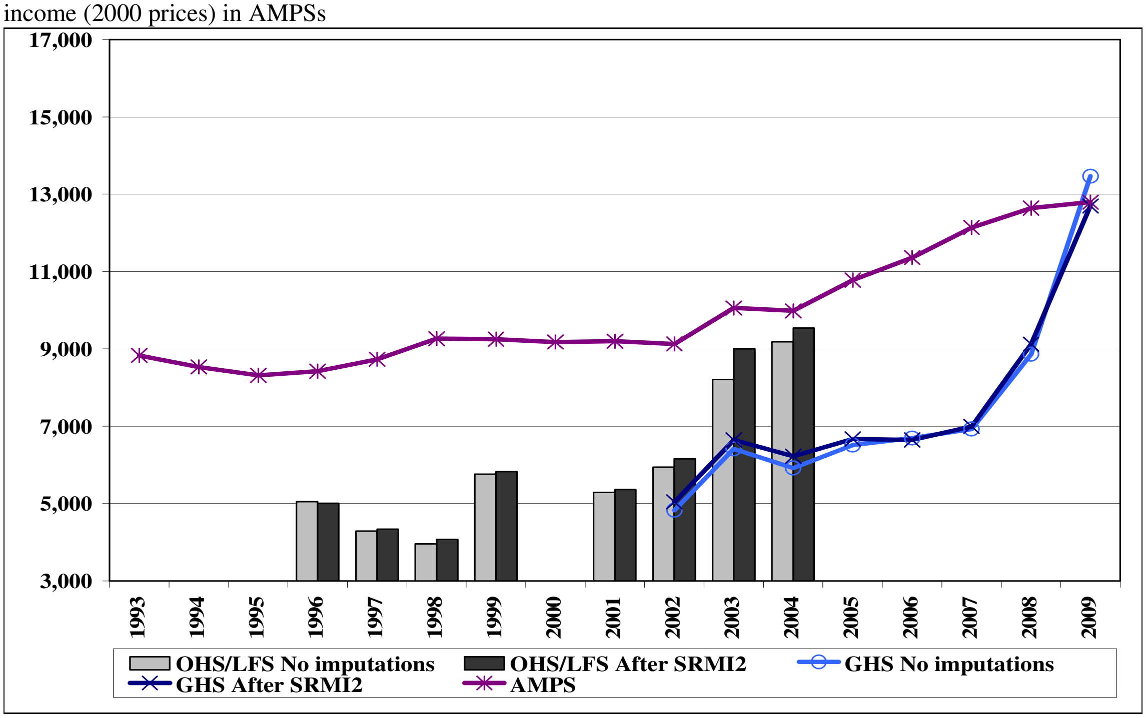


Figure 5.10: Mean annual per capita expenditure (2000 prices) in the OHSs, LFSs, GHSs and mean per capita income (2000 prices) in AMPSs



5.5 Comparison with national accounts income data

Section 3.9 discussed the pros and cons of adjusting the survey distribution rightwards in line with the national accounts mean income. In this section, the total income, expenditure or consumption amounts derived from different surveys in different years are compared with the national accounts income data, so as to see if the surveys did not capture income / expenditure / consumption information precisely. In addition, the income and expenditure variables in the surveys are compared with one another to find out if the differences, if any, have to do with the factors discussed in Chapter 3.

First, Table 5.16 shows the derivation of the total income in national accounts. Due to changes in the categorization of items in the Quarterly Bulletin of the Reserve Bank since 2006, the formula to calculate total income has changed[95].

Table 5.16: Derivation of the total income in national accounts

Old method: Until 2005	
Remuneration	Compensation of employees (6240)
Transfers	Transfers from general government (6257)
Residuals	Property income (6241) + Current transfer from enterprise (6231) + Transfer from rest of the world (6243)
New method: Since 2006	
Remuneration	Compensation of employees (6240)
Transfers	Gross operating surplus/mixed income (6826) + Property income received (6827) – Property income paid (6832) – Consumption of fixed capital (6849)
Residuals	Social benefits received (6836) + Other current transfers received (6837) – Social contributions paid (6840) – Other current transfers paid (6841) + Adjustment for change in net equity of households in pension funds reserves (6845)

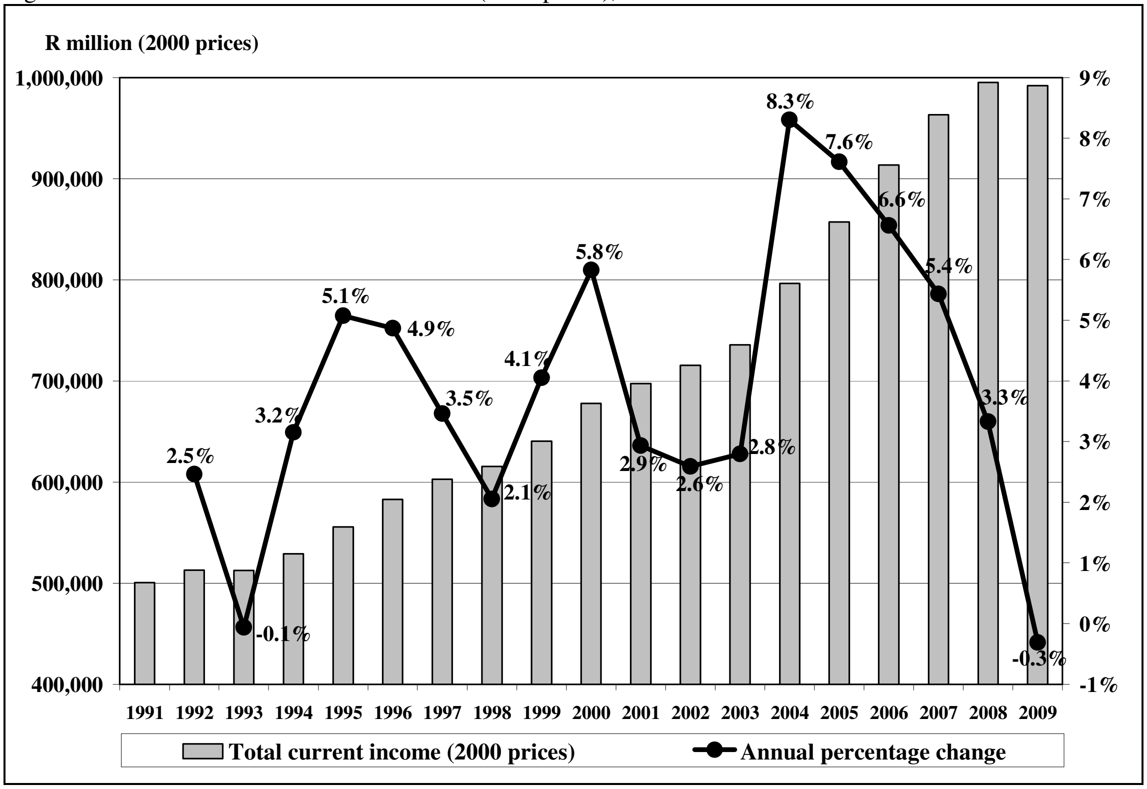
Note: The code of each item in the South African Reserve Bank Quarterly Bulletin is shown in brackets.

Figure 5.11 shows that total income (in 2000 prices) almost doubled between 1992 and 2008. Moreover, the total income showed a positive growth in all years, except between 1992 and

[95] In the national accounts section of the Bulletin's statistical tables before 2006, the current income of the households was clearly shown as the sum of compensation of employees, transfers from general government, property income, current transfer from enterprise and transfer from rest of the world. However, since 2006, the categorization of the items has changed drastically, and the current income could only be approximated using the items shown in the last three rows of Table 5.16. These items are under the statistical table "Production, distribution and accumulation accounts of South Africa – Households and non-profit institutions serving households" in the Bulletin.

1993 (-0.1%) and between 2008 and 2009 (-0.3%). The decrease in these two years could be attributed to the political uncertainty and the impact of the global recession on the South African economy respectively.

Figure 5.11: Total income in national accounts (2000 prices), 1991-2009



Note: The annual percentage change of total income between 1991 and 1992 was 2.5%, while it was -0.1% between 1992 and 1993, and so forth.

Table 5.17 as well as Figures 5.12 and 5.13 show the total income, expenditure or consumption in each survey as percentage of the national accounts total income in the same year. First, looking at the two censuses and CS 2007, it can be seen that CS 2007 is the survey that captured total income the best. In addition, as expected, after dealing with households with zero or unspecified income by SRMI1 and SRMI2, total income became higher in all three surveys.

With regard to the IESs, the 1995 survey best captured total income and expenditure. Under the STC approach, these amounts were equal to nearly 95% of the national accounts income amount. One notable finding from Table 5.17 is that the IES total income / expenditure / consumption experienced a sharp decline between 1995 and 2000, which contradicted the upward trend in the national accounts total income as seen in Figure 5.11.

Table 5.17: Comparison of annual total income/expenditure/consumption in various surveys with annual total income in the national accounts in the same year

Survey	Variable	Year	Amount (R million) (2000 prices)	As % of total income in the national accounts
Census/CS	Total income – without any imputations involved	1996	294 475	50.5%
		2001	366 341	52.5%
		2007	629 421	68.9%
	Total income – After SRMI1	1996	339 993	58.3%
		2001	470 360	67.4%
		2007	776 476	85.0%
	Total income – After SRMI2	1996	350 345	60.1%
		2001	506 896	72.7%
		2007	782 283	85.6%
IES	Total income – STC	1995	527 850	95.0%
		2000	460 572	71.9%
		2005/2006	659 229	72.2%
	Total expenditure – STC	1995	519 549	93.5%
		2000	458 867	71.7%
		2005/2006	751 153	82.2%
	Total income - COICOP	1995	495 411	89.2%
		2000	441 795	69.0%
		2005/2006	705 713	77.3%
	Total consumption - COICOP	1995	365 935	65.9%
		2000	324 026	47.8%
		2005/2006	531 386	58.2%
OHS	Total expenditure – No imputations	1996	190 111	32.6%
		1997	172 608	28.6%
		1998	151 399	24.6%
		1999	229 693	35.9%
	Total income – No imputations	1999	607 350	94.9%
	Total expenditure – After SRMI2	1996	195 845	33.6%
		1997	183 153	30.4%
		1998	161 717	26.3%
		1999	252 422	39.4%
	Total income – After SRMI2	1999	746 173	116.5%
LFS	Total expenditure – No imputations	2001	230 514	33.1%
		2002	264 065	36.9%
		2003	370 790	50.4%
		2004	417 062	52.4%
	Total expenditure – After SRMI2	2001	241 690	34.7%
		2002	280 567	39.2%
		2003	414 435	56.3%
		2004	443 144	55.6%
GHS	Total expenditure – No imputations	2002	212 412	29.7%
		2003	287 893	39.1%
		2004	267 470	33.6%
		2005	299 400	34.9%
		2006	312 736	34.2%
		2007	326 385	33.9%
		2008	461 528	46.7%
		2009	606 047	61.1%

Table 5.17: Continued

Survey	Variable	Year	Amount (R million) (2000 prices)	As % of total income in the national accounts
GHS	Total expenditure – After SRMI2	2002	229 177	32.0%
		2003	308 977	42.0%
		2004	289 165	36.3%
		2005	312 468	36.5%
		2006	314 442	34.4%
		2007	334 237	34.7%
		2008	486 045	49.2%
		2009	612 482	61.7%
PSLSD	Total income	1993	334 531	65.3%
	Total expenditure	1993	297 679	58.1%
NIDS	Total income	2008	627 815	63.1%
	Total expenditure	2008	546 682	54.9%
AMPS	Total income	1993	336 394	65.6%
		1994	330 381	62.5%
		1995	333 057	59.9%
		1996	349 167	59.9%
		1997	347 982	57.7%
		1998	361 044	58.7%
		1999	360 573	56.3%
		2000	404 993	59.8%
		2001	406 077	58.2%
		2002	403 762	56.4%
		2003	444 193	60.4%
		2004	450 696	56.6%
		2005#	485 001	56.6%
		2006#	502 572	55.0%
		2007#	552 266	57.3%
		2008#	629 142	63.2%
		2009#	589 559	59.4%

Originally, the AMPS income showed a rapid 22.4% increase between 2004 and 2005 (the 2005 figure being R551 433 million. Thus, it was decided to use the 2004-2005 national accounts income growth rate (7.61%) to derive the 'adjusted' AMPS 2005 income. Furthermore, the annual income growth rates of AMPS data were used as they are to derive the 'adjusted' AMPS 2006-2009 income.

In other words:

- "Adjusted" 2005 AMPS income = 450 696 × (1 + 7.61%) = 485 001
- "Adjusted" 2006 AMPS income = 485 001 × (1 + 3.62%) = 502 572
- "Adjusted" 2007 AMPS income = 502 572 × (1 + 9.89%) = 552 266
- "Adjusted" 2008 AMPS income = 552 266 × (1 + 13.92%) = 629 142
- "Adjusted" 2009 AMPS income = 629 142 × (1 – 6.29%) = 589 559

Figure 5.12: Total income, consumption or expenditure as percentage of national accounts' total income in the IESs, PSLSD, OHS 1999 (Income), NIDS, censuses and CS 2007

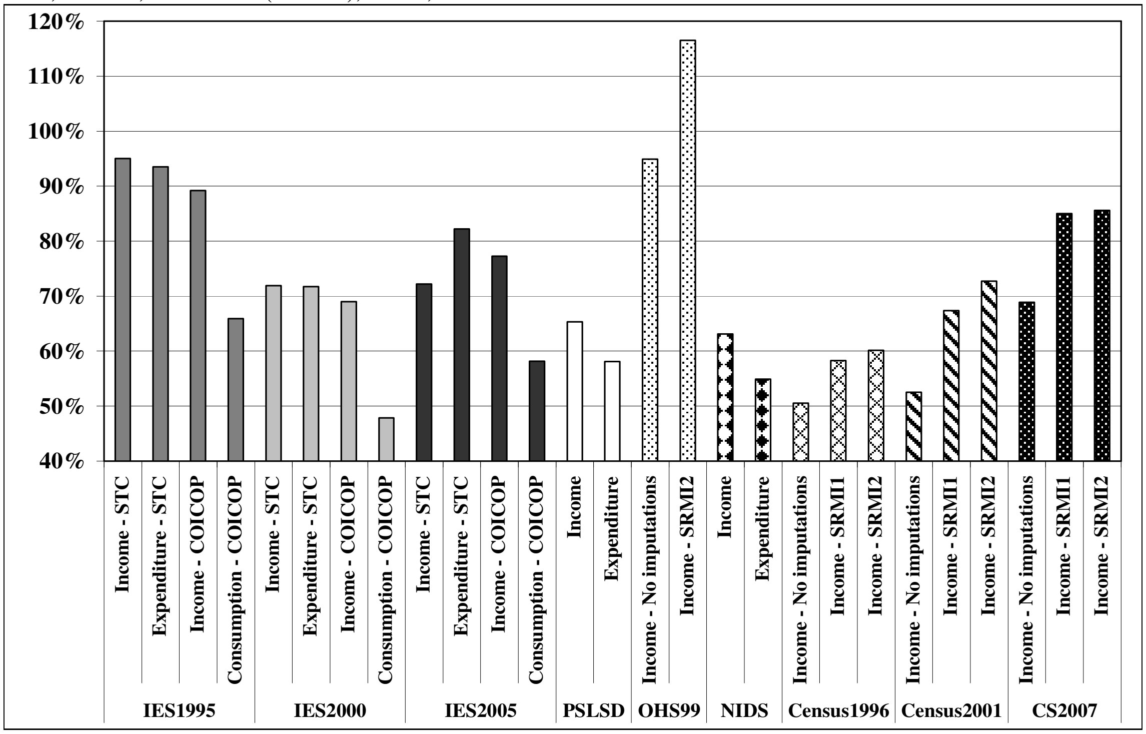
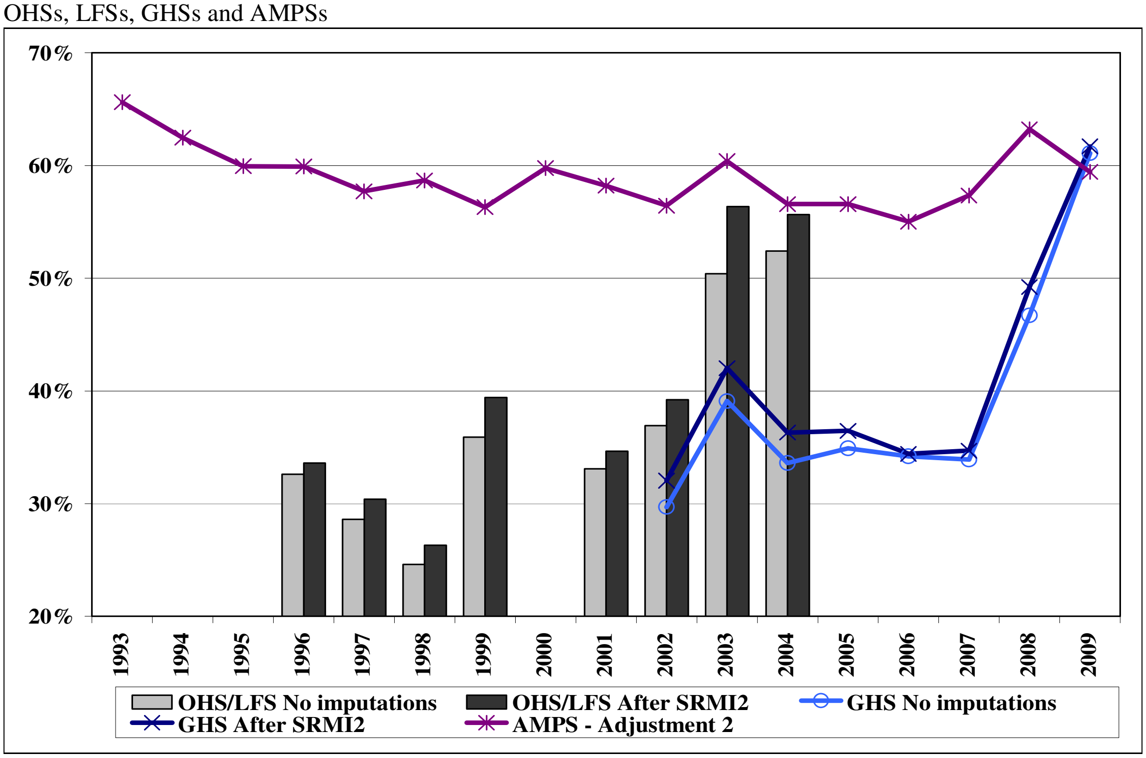


Figure 5.13: Total income, consumption or expenditure as percentage of national accounts' total income in the OHSs, LFSs, GHSs and AMPSs

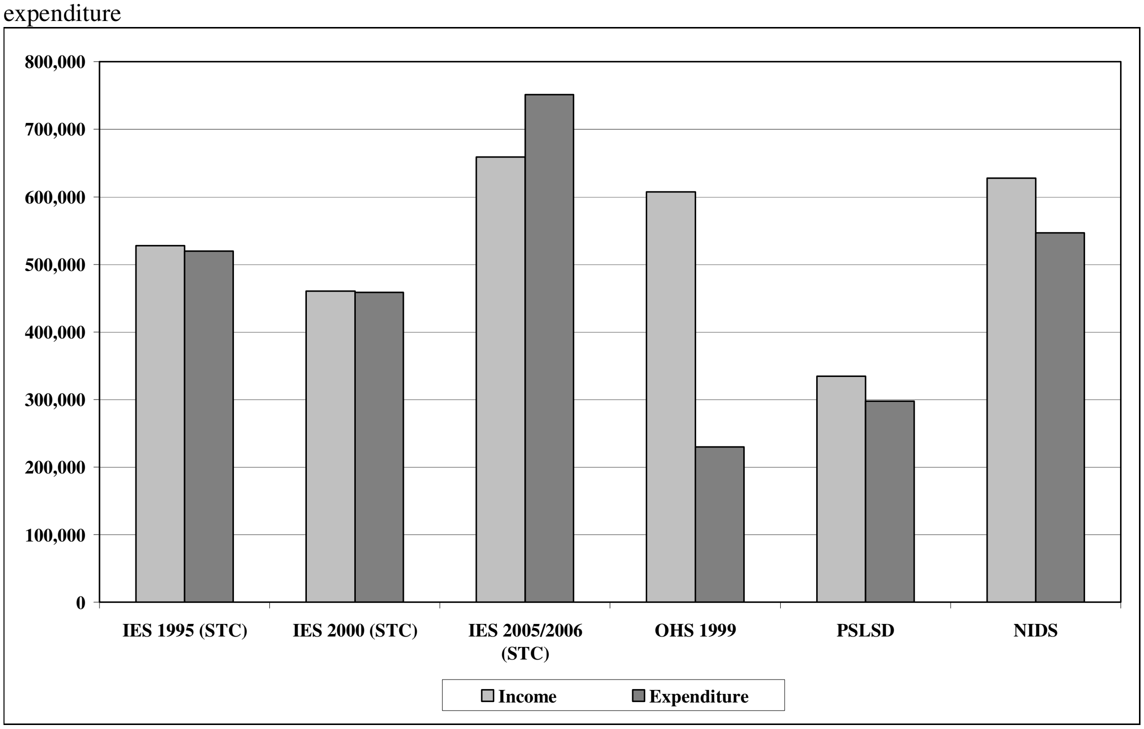


Total expenditure was seriously under-captured in the OHSs, LFSs and GHSs, as it only amounted to 30%-50% of the national income (except GHS 2009), and this proportion only increased slightly even after SRMI2 was applied. It is possible that the fatigue factor as discussed in Section 3.12 could play a role to account for the very low amounts captured in the OHSs/LFSs/GHSs, since the question on household expenditure was only asked at the end of the survey. Nonetheless, OHS 1999 captured income extremely well, as total household income amounted to 94.9% of the 1999 national income without imputation and 116.5% (i.e., exceeding the national income of 1999) after SRMI2.

As far as the surveys conducted by institutions other than Stats SA are concerned, in both PSLSD and NIDS, total expenditure was under-captured more seriously than total income, when compared with national accounts' total income. Finally, in almost all AMPSs, total income was approximately 60% of national income.

Turning the attention to the comparison of income and expenditure across the surveys, Figure 5.14 shows that, for surveys that captured both of them, income was greater than expenditure in all surveys with the exception of IES 2000, contradicting the general argument as discussed in Section 3.2 that expenditure is captured better in poor, developing countries.

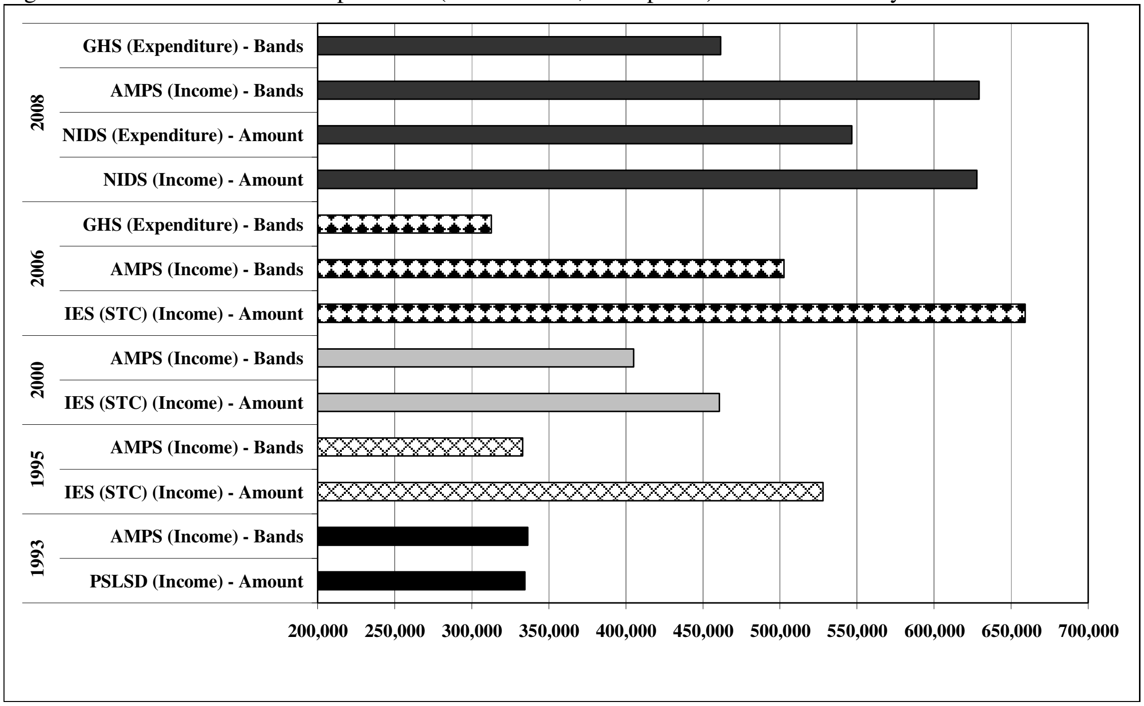
Figure 5.14: Total income and expenditure (Rand million, 2000 prices) of surveys that collected both income and expenditure



With regard to whether the income and expenditure information is captured better if the respondents are asked to declare the actual continuous amount instead of the relevant interval,

Figure 5.15 shows that there is no such indication. For instance, the IES 1995 total income (respondents were asked to declare the information in exact amounts) was greater than the 1995 AMPS total income (the interval method was used), while in 2008, the AMPS income (interval method was adopted) was greater than both the NIDS income and expenditure (actual continuous amounts were declared), but the GHS expenditure (interval method was used) was the lowest.

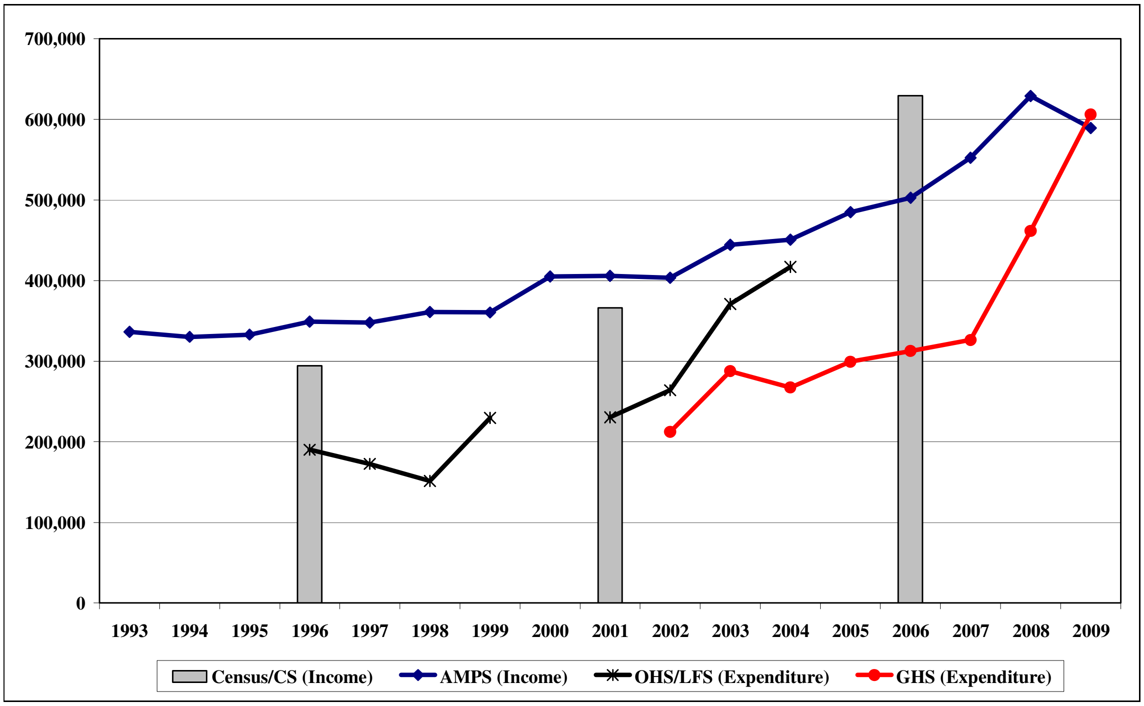
Figure 5.15: Total income and expenditure (Rand million, 2000 prices) of selected surveys



As far as the relationship between income / expenditure and the number of intervals as well as the width of each interval is concerned (for surveys that adopted the interval method to capture these information), Figure 5.16 shows that total expenditure in OHSs, LFSs and GHSs (with very few intervals for the respondents to choose from) was clearly lower. On the other hand, total income in Census 1996 and 2001 (with very wide intervals in the higher-income categories) was lower than the total income captured in the same years in AMPS, which has more income intervals and narrower width in each interval, but the opposite happened when comparing CS 2007 with AMPS 2007. Finally, the AMPS income was always bigger than the GHS expenditure from the same year, except in 2009.

The impact of the number of intervals and the width of each interval on poverty and inequality estimates is a research issue that has never been investigated thoroughly in the South African studies, and Chapter 6 will come back to this issue.

Figure 5.16: Total income and expenditure (Rand million, 2000 prices) of surveys that adopted the interval method



Finally, the IES expenditure (STC approach) and consumption (COICOP approach) variables are analysed in greater detail in Tables 5.18 and 5.19 to find out whether the diary method adopted in IES 2005/2006 has any influence on the preciseness of these estimates. It can be seen that food expenditure, which was captured entirely by the diary method in IES 2005/2006, surprisingly decreased when compared with IES 2000. Food expenditure as proportion of total expenditure also abruptly declined between these two surveys, as shown in Figure 5.17. This result contradicts what was found in the GHSs, as the proportion of households reporting they never experienced adult hunger and child hunger[96] in the past 12 months at the time of the survey increased continuously[97].

It is possible that the diary method resulted in the under-estimation of food expenditure (due to reasons like first-day effect, illiteracy of respondents, as discussed in Section 3.3). Nonetheless, it is also possible that the recall method adopted in IES 1995 and IES 2000 resulted in an over-estimation of food expenditure due to factors like recall bias and telescoping (see Section 3.3), while the IES 2005/2006 food expenditure estimate could be more reliable.

---

[96] Only households with children aged 0-17 years were included.

[97] The proportion of households never experiencing adult hunger increased from 69.25% in GHS 2002 to 83.42% in GHS 2008, while the proportion of households never experiencing child hunger increased from 69.20% to 82.44% between these two surveys.

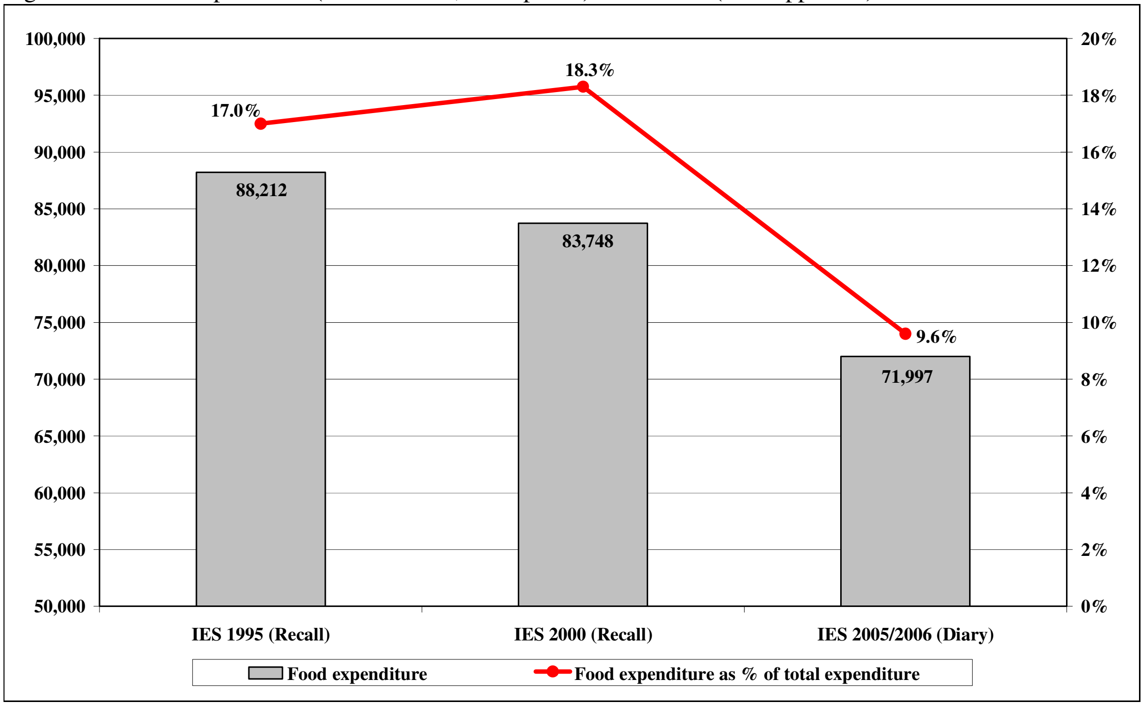
Table 5.18: Total expenditure (Rand million, 2000 prices) in each category using STC approach in all three IESs

Expenditure category	IES 1995		IES 2000		IES 2005/2006	
(1) Housing	76 084	14.6%	78 656	17.1%	118 512	15.8%
(2) Domestic workers	7 251	1.4%	11 703	2.6%	10 615	1.4%
(3) Food	88 212	17.0%	83 748	18.3%	71 997	9.6%
(4) Beverages	8 433	1.6%	9 781	2.1%	7 616	1.0%
(5) Cigarettes and smokers' requisites	4 343	0.8%	4 530	1.0%	3 680	0.5%
(6) Personal care	11 354	2.2%	14 242	3.1%	6 603	0.9%
(7) Other household consumer goods	6 534	1.3%	4 821	1.1%	4 229	0.6%
(8) Household services	1 612	0.3%	446	0.1%	323	0.0%
(9) Household fuel	2 726	0.5%	4 087	0.9%	3 386	0.5%
(10) Clothing and footwear	23 440	4.5%	16 981	3.7%	26 304	3.5%
(11) Furniture/Equipment	18 923	3.6%	10 602	2.3%	21 234	2.8%
(12) Health services	18 678	3.6%	16 937	3.7%	29 978	4.0%
(13) Transport	48 988	9.4%	46 986	10.2%	110 498	14.7%
(14) Computer and telecommunication equipment	1 502	0.3%	3 071	0.7%	4 655	0.6%
(15) Communication for household purposes	10 907	2.1%	9 613	2.1%	16 414	2.2%
(16) Education	8 822	1.7%	13 160	2.9%	18 558	2.5%
(17) Reading matter and stationery	2 298	0.4%	3 109	0.7%	2 678	0.4%
(18) Recreation, entertainment and sports	6 457	1.2%	7 147	1.6%	15 258	2.0%
(19) Miscellaneous expenditure	166 270	32.0%	110 123	24.0%	274 949	36.6%
(20) Expenditure on own harvest/livestock	6 714	1.3%	9 123	2.0%	3 667	0.5%
Total household annual expenditure	***519 549***	***100.0%***	***458 867***	***100.0%***	***751 153***	***100.0%***

Table 5.19: Total consumption (Rand million, 2000 prices) in each category using COICOP approach in all three IESs

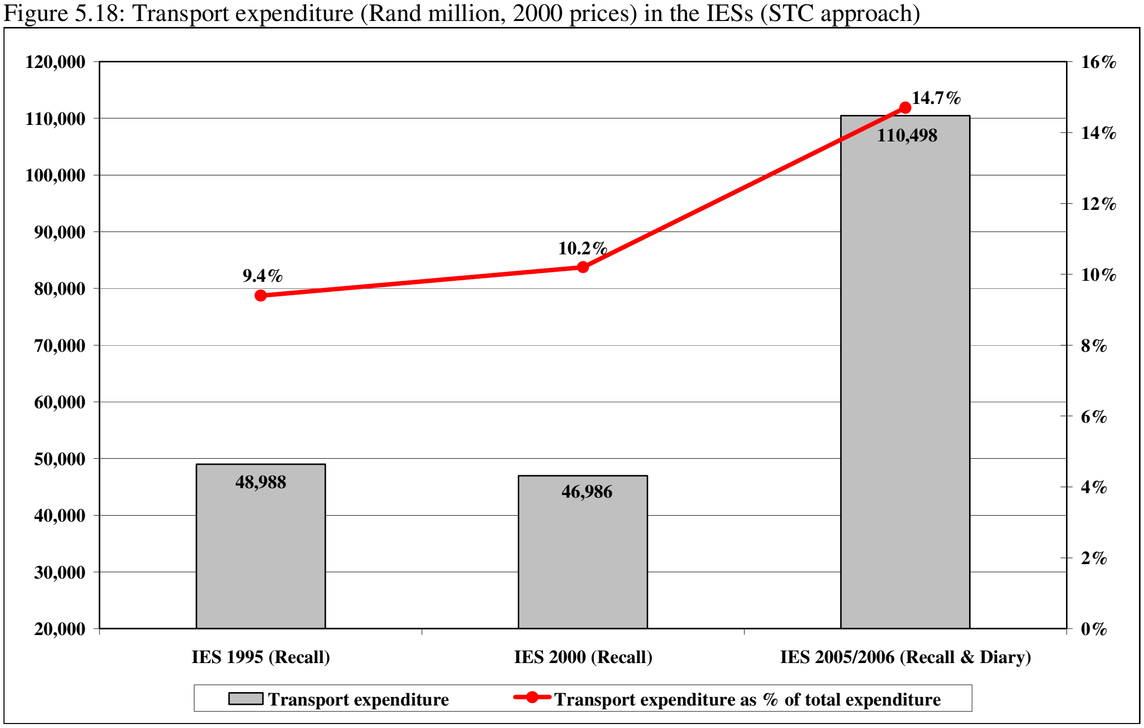
Consumption category	IES 1995		IES 2000		IES 2005/2006	
(1) Food and non-alcoholic beverages	90 809	24.82%	88 771	27.4%	76 771	14.4%
(2) Alcoholic beverages, tobacco and narcotics	8 061	2.20%	8 179	2.5%	6 135	1.2%
(3) Clothing and footwear	23 508	6.42%	17 084	5.3%	26 300	4.9%
(4) Housing, water, electricity, gas and other fuels	40 432	11.05%	43 716	13.5%	125 213	23.6%
(5) Furnishings, household equipment and maintenance of house	34 538	9.44%	27 787	8.6%	36 608	6.9%
(6) Health	3 869	1.06%	4 800	1.5%	8 834	1.7%
(7) Transport	42 780	11.69%	42 663	13.2%	105 801	19.9%
(8) Communication	11 330	3.10%	10 764	3.3%	18 638	3.5%
(9) Recreation and culture	11 331	3.10%	14 387	4.4%	24 455	4.6%
(10) Education	5 600	1.53%	9 009	2.8%	12 825	2.4%
(11) Restaurants and hotels	6 941	1.90%	11 445	3.5%	11 668	2.2%
(12) Miscellaneous goods and services	38 934	10.64%	45 421	14.0%	76 514	14.4%
(13) Other unclassified expenses	47 803	13.06%	0	0.0%	1 623	0.3%
Total consumption	***365 935***	***100.00%***	***324 026***	***100.0%***	***531 386***	***100.0%***

Figure 5.17: Food expenditure (Rand million, 2000 prices) in the IESs (STC approach)



In contrast, regardless of whether the STC or COICOP approach was adopted, transport expenditure was much higher in IES 2005/2006 (see Tables 5.18 and 5.19 as well as Figure 5.18 below). It is possible that the use of the diary method to complement the recall method (see Table 2.14) resulted in a better capture of transport expenditure in this survey.

Figure 5.18: Transport expenditure (Rand million, 2000 prices) in the IESs (STC approach)



5.6 Chapter summary

Chapter 5 began by reviewing the poverty definitions, before discussing the possible money-metric and non-money-metric indicators that could be used to measure poverty. It then reviewed the various approaches to define the poverty line, before discussing in detail the three FGT poverty indices. Next, inequality concepts and measurements were discussed, with particular focus on the Lorenz curve, Gini coefficient, and the two Theil indices.

The chapter moved on to apply the SRMI methodology as discussed in Chapter 3 to deal with households reporting zero or unspecified amounts, before per capita income / expenditure / consumption variables in 2000 prices were derived in the surveys under study. These variables were then compared with the national accounts income of the same year, so as to find out if some surveys seriously under-captured income or expenditure. It was found that this took place in all survey income or expenditure variables, except the OHS 1999 income variable. Furthermore, as discussed in Chapter 3, in developing countries, expenditure is likely to be captured better than income, but the results of the analyses in Section 5.5 did not indicate that the survey expenditure was always greater than the survey income (or the survey expenditure was closer to the national accounts income), when looking at surveys that captured both income and expenditure. Finally, the income and expenditure estimates were also compared across the surveys, and it was found that the other factors discussed in Chapter 3 might have played a role to cause these estimates to be different across the surveys.

In Chapter 6, poverty and inequality trends will be investigated. The money-metric per capita income / expenditure / consumption variables will be used. In addition, the three objective absolute poverty lines using the cost of basic needs method as derived by Woolard and Leibbrandt (2006) will be used to measure poverty.

CHAPTER SIX: POVERTY AND INEQUALITY TRENDS IN SOUTH AFRICA SINCE THE TRANSITION

6.1 Introduction

This chapter examines the poverty and inequality trends in South Africa since the transition by using the real per capita income, expenditure and consumption variables derived from various surveys as discussed in Chapter 5. Chapter 6 begins by providing a literature review of the results of recent studies on the poverty and income distribution using the money-metric approach in Section 6.2. This is followed by Sections 6.3 and 6.4, which investigate poverty and inequality trends respectively. With regard to the poverty trends, the three poverty lines proposed by Woolard and Leibbrandt (2006) are used. The impact of the inclusion of the imputed rent variable in total income on the poverty and inequality estimates is briefly looked at in Section 6.5.

As explained in Chapter 3, SALDRU was concerned that the 'one-shot', single estimate of income and expenditure derived in NIDS was an under-estimation, and consequently decided to use the aggregation approach to derive the income and expenditure variables. These latter variables were used by SALDRU in the poverty and inequality analyses. The focus of Section 6.6 is on investigating whether the poverty and inequality estimates would differ significantly using the variables derived in each approach. Section 6.7 analyses what would have happened to income poverty and inequality estimates in the IESs and NIDS, had the respondents been asked to declare the relevant income category instead of having the exact amounts derived, and were given the categories used in Census, GHS and AMPS respectively, so as to investigate the impact of the number and width of intervals on poverty and inequality estimates and trends. Furthermore, the poverty and inequality trends after adjusting the survey data in line with the national accounts figures and after applying the cross entropy re-weighting approach are looked at in Sections 6.8 and 6.9 respectively. Section 6.10 concludes the chapter.

6.2 Literature review of recent studies on poverty and inequality trends

In this section, the results of recent studies that adopted the money-metric approach to investigate poverty and inequality trends are discussed. The majority of these studies analysed the IES or census data to derive these trends, while the AMPS data have become an alternative data source for these analyses by a team of researchers at Stellenbosch University. Studies not discussed here are those that only used one household survey to derive a snapshot poverty and inequality profile of South Africa (e.g., Klasen 1997 & 2000; Woolard and Leibbrandt 1999; Woolard 2001; Armstrong, Lekezwa and Siebrits 2008; Argent, Finn, Leibbrandt and Woolard 2009; Finn, Leibbrandt and Woolard 2009; Posel & Rogan 2009; Rogan 2010), as well as those that used more than one survey to derive poverty and inequality trends using non-money-metric approaches (e.g., Bhorat, Poswell and Naidoo 2004; Bhorat, Naidoo and Van der Westhuizen 2006; Bhorat, Van der Westhuizen and Goga 2007).

6.2.1 Review of studies that used the censuses

Leibbrandt, Poswell, Naidoo and Welch (2006) compared Census 1996 with Census 2001. With regard to Census 2001, they used the personal income variable before hot deck imputation by Stats SA. Furthermore, in both censuses, the incomes of people below 15 years with positive personal incomes were reset as zero, before the household income was derived. Households with zero or unspecified household income were excluded, before per capita income was derived. Using the poverty lines of R250 per month (1996 prices) and US$2 per day, they found that the poverty headcount ratio increased across the two censuses (from 0.50 to 0.55 for the R250 per month poverty line and from 0.26 to 0.28 for the US$2 per day poverty line). In addition, the Gini coefficient increased from 0.68 to 0.73. This increase was caused by the more rapid increase of real household income in the top income deciles in Census 2001. Furthermore, within-race inequality's share of total inequality increased from 57% to 60% across the two censuses. Finally, after including households with zero income for the analyses, they found that the poverty headcount ratio increased from 0.59 to 0.65 between the two censuses using the R250 per month poverty line, and from 0.40 to 0.44 using the US$2 per day poverty line. Income inequality also worsened, as indicated by the increase of the Gini coefficient from 0.74 to 0.79.

Two problems of the Leibbrandt et al. study were that the incorrect 1996 household income variable derived by Stats SA (see Table 2.4) was used, and households with zero or unspecified household income were simply excluded. This could affect the validity of the poverty and inequality estimates. These two problems were addressed in this dissertation by using the correct 1996 income variable and imputing the income of the abovementioned households by SRMI. Sections 6.3 and 6.4 will examine how the poverty and inequality estimates and trends would be influenced as a result of addressing these issues.

In the study by Ardington, Lam, Leibbrandt and Welch (2005), some dubious zero personal income values (e.g., employed adults reporting zero income) were set to missing, before SRMI was conducted at household level to derive the imputed household income category. Next, the interval data was made continuous, but the commonly used midpoint-Pareto method was not used by the authors. Instead, the cumulative income distribution of the IES 2000 personal incomes in 2001 prices was generated for each census income band. Next, random probabilities were generated for each person in the census, and each individual was assigned an income amount such that "the cumulative probability of observing such a value from the empirical distribution was greater than or equal to the generated probability" (Ardington et al. 2005: 18). The personal income amounts of the members from a household were then added, before being divided by household size to derive the per capita income variable. A similar exercise was done in the IES 1995 data to make the Census 1996 personal income interval data continuous.

Using such revised per capita income variables to estimate poverty and inequality in the two censuses, the authors found that the Gini coefficient increased from 0.74 to 0.82 across the two surveys. In addition, the poverty headcount ratios increased from 0.38 to 0.42 using the R124 per month (2001 prices) poverty line, and from 0.60 to 0.68 using the R400 per month (2001 prices) poverty line. Like Leibbrandt et al. (2006), Ardington et al. (2005) also used the incorrect household income variable of Census 1996 in their study. To conclude, the three studies above all found that the poverty headcount ratio and the Gini coefficient increased between Census 1996 and Census 2001.

6.2.2 Review of studies that used the IESs

Stats SA (2002e) compared the IES 1995 and IES 2000 per capita income and expenditure derived by the STC method, in order to paint a picture of how income was distributed in South Africa by using five income quintiles, and also to examine expenditure in households by expenditure quintiles. Stats SA founded that both household and per capita income and expenditure decreased over time. The Gini coefficient at household level increased from 0.56 to 0.57 if per capita income was used. Furthermore, in both surveys, the poorest 10% of households received as little as 0.5% of all household income, while the richest 10% earned approximately 45% of all income in both surveys. Poor households spent proportionately more on food but saved less, compared with the more affluent ones.

Leibbrandt, Levinsohn and McCrary (2005) focused on the personal income of individuals aged at least 18 years. They found that personal income decreased by 40% across the two IESs. This resulted in a sharp increase of poverty between the two years, according to their calculations. Such a dramatic decline in income contradicted the upward trends in the current household income as captured in the national accounts. They suggested that the main reason for the strong increase in poverty was a strong decline in return to endowments such as education. The econometric analyses presented evidence of decreasing returns to education amongst the black population, compared with the rising returns for whites.

Hoogeveen and Özler (2006) investigated poverty and inequality by using the IES 1995 and IES 2000 per capita expenditure data (adopting the STC categorization approach). They applied two poverty lines – the international US$2 per day (which is equivalent to R174 per month in 2000 prices) poverty line and a cost-of-basic-needs poverty line of R322 per month (also in 2000 prices) to investigate the extent of extreme poverty and moderate poverty respectively. The results indicated that extreme poverty increased slightly between the two surveys (the poverty headcount ratio increasing from 0.32 to 0.34), and this was attributed to rising poverty among the black population in Gauteng and rural provinces containing the former homelands. In addition, moderate poverty remained unchanged (with the poverty headcount ratio remaining at 0.58). The Gini coefficient increased slightly and inequality worsened most seriously amongst the black population. The white population was the only race group experiencing a decline in Gini coefficient. Finally, the Theil-L index showed that the share of within-race inequality to total inequality increased from 61.7% in 1995 to 66.8% in 2000. Özler (2007) investigated the IES expenditure data by excluding certain spending categories[98] when deriving a revised per capita expenditure. However, the poverty and inequality trends derived from this revised variable were very similar to the results by Hoogeveen and Özler (2006).

Van der Berg & Louw (2004) as well as Pauw and Mncube (2007) were concerned that the rapid decline in both household income between the two aforementioned IESs was inconsistent with the increase of the current household income as captured in the national accounts. Hence, Van der Berg and Louw calculated mean incomes by race using national accounts and other sources of data, before applying these income values to the intra-group

[98] The expenditure categories excluded were: health, water, firewood and dung, imputed value of household durables, food consumption from home production, lobola/dowry, funerals, religious or traditional ceremonies, gambling, as well as lumpy expenditures such as appliances, vehicles, furniture, sound and video equipments and so on.

distributions of income contained in these two IESs. Using a poverty line of R250 per month in 2000 prices, the poverty headcount ratio declined slightly across the two surveys, but the number of people living in poverty increased due to population growth. The Gini coefficients increased slightly in all race groups, but the Gini coefficient of the whole population was not presented.

Pauw and Mncube (2007) first imputed the food and tax expenditures in IES 2000 so as to replace unexpected missing or zero values or cases of obvious under-reporting. They also assumed that for each household in both surveys, the larger of total income and total expenditure was the correct measure, which was then used to derive the per capita variable for the poverty and inequality analyses. Three poverty lines (US $1 a day, US $2 a day and the lower bound R322 per month in 2000 prices) were used. The poverty headcount ratio increased moderately between the two surveys in all three cases, while the Gini coefficient increased from 0.62 to 0.66.

With regard to the studies that compared IES 1995 with IES 2005/2006, Bhorat and Van der Westhuizen (2008) used the per capita consumption variable derived using the COICOP approach to investigate the critical interactions between economic growth, poverty and inequality. Their results showed that the poverty headcount ratio decreased between 1995 and 2005 for the two poverty lines chosen (R174 and R322 per month in 2000 prices). The Gini coefficient increased from 0.64 to 0.69. Surprisingly, the Theil-T index showed that within-race inequality as proportion of total inequality decreased from 53.2% to 50.4% across the two surveys. Furthermore, the growth incidence curves showed that although growth in per capita consumption was pro-poor in an absolute sense as all individuals across the distribution experienced positive growth, it was not pro-poor in a relative sense, because the individuals in the top 10% of the distribution enjoyed the greatest growth rates of all. Finally, the poor at the bottom of the distribution experienced growth in consumption mainly due to the rapid expansion of the social security system.

Bhorat, Van der Westhuizen and Jacobs (2009) also compared IES 1995 with IES 2005/2006 by using the per capita income variable derived using the COICOP approach. They found that the Gini coefficient increased from 0.64 to 0.72 across the two surveys. The inequality decomposition of the Theil-T index found that the share of within-race inequality to total inequality decreased slightly from 57.4% to 55.6%, consistent with what was found by Bhorat and Van der Westhuizen (2008). Furthermore, an income decomposition by income source was conducted across the two surveys, and it was found that wage income's contribution to income inequality became more important across the two surveys. It was argued that this was caused by the increase of earnings of highly-skilled workers at the top end of the distribution.

The study by Yu (2008) was the first attempt to use all three post-transition IESs to look at the poverty and inequality trends. His study did not investigate the trends in too much detail (i.e., by race, gender, etc.) but only provided preliminary findings on the aforementioned trends. Using the per capita income variables derived by the STC approach across all three surveys and the poverty line value of R322 per month (2000 prices), the poverty headcount ratio increased between IES 1995 and IES 2000 (from 0.44 to 0.56), before decreasing to just below 0.50 in IES 2005/2006. A similar result was observed when using the per capita income variable derived by the COICOP approach in all surveys. With regard to the inequality trends, the Gini coefficient increased between 1995 and 2000, before a negligible increase was observed in the most recent IES. These findings were observed for both income variables.

To conclude, the studies that used the IES data found that the poverty headcount ratio and the Gini coefficient increased between 1995 and 2000. However, with regard to the latter variable, some studies found that the Gini coefficient increased moderately while others found that it increased marginally. In addition, studies that also took IES 2005/2006 data into consideration found that the poverty headcount ratio decreased between the 2000 and 2005/2006 IESs, but the Gini coefficient stabilised.

6.2.3 Review of studies that used the OHSs and LFSs

Only two studies used the OHS/LFS data to investigate the poverty and inequality trends since the transition. In Meth and Dias (2004), the OHS 1999 and LFS 2002b household expenditure information was used to derive per adult equivalent expenditure. Using a poverty line of R467 per adult equivalent per month in 2000 prices, the authors found that the number of poor increased by at least 2 million to about 4.5 million in 2002. The study did not investigate the inequality trends. In the study by Vermaark (2005), OHS 1995, OHS 1997, LFS 2001b and LFS 2003b earnings from main job[99] data were used to derive per adult equivalent earned income. The poverty headcount ratio increased continuously for the two poverty lines chosen (R152 and R405 per adult equivalent per month in 2000 prices). In

[99] Earnings from the main job is only one of the sources of income. Income could also be earned from the secondary job, interest from investment, social grant income received from the government. Hence, earnings from the main job is lower than total income.

addition, the Gini coefficient increased between 1995 and 1997, before it dropped to the 1995 level in 2001 and 2003. Finally, the inequality decomposition of the Theil-T index showed that within-race inequality's contribution to total inequality of per capita earnings increased rapidly between 1995 and 1997, and decreased between 1997 and 2001, before stabilising in 2003.

6.2.4 Review of studies that used the AMPSs

As mentioned in Chapter 1, the AMPS data became a newly used data for poverty and inequality trends analyses by the team of Stellenbosch University researchers. Four studies were done so far using this alternative data source. First, Van der Berg, Louw and Yu (2008) used the 1993-2004 AMPS and chose a poverty line of R250 per month (2000 prices). They found that the poverty headcount ratio increased continuously in 1993-1996, followed by a period of approximate stability until the turn of the century (although the number of poor increased during this period as a result of population growth), and then by a strong reduction after 2001, with the number of poor even decreasing substantially despite population growth. The increase in poverty rate in the 1990s was attributed to a combination of sluggish economic growth and poor labour market prospects. In contrast, the decline in poverty in the 2000s was driven by faster economic growth, improving labour market prospects and expansion of social grant expenditure. The Gini coefficient increased slightly between 1993 and 2000, before stabilising in the 2000-2004 period. Finally, the Theil-L and Theil-T inequality measures both showed that the contribution of the within-race component to total inequality increased continuously between 1993 and 2004. Van der Berg, Louw and Du Toit (2008) conducted a similar study by including the 2005 and 2006 AMPS data, and had the same findings.

The two studies by Van der Berg, Burger, Burger, Louw and Yu (2005 & 2009) also used the AMPS income data, but the distributional estimates were adjusted to be consistent with the national accounts series for aggregate household income. In their 2005 study, using the 1993-2004 data and once again the R250/month (2000 prices) poverty line, the poverty headcount ratio increased between 1993 and 2000, before a more rapid downward trend was observed between 2000 and 2004. Furthermore, the number of poor increased from 16.2 to 18.5 million between 1993 and 2000, before dropping by 3.2 million to 15.4 million in 2004. With regard to inequality trends, the Gini coefficient increased slightly for all race groups during the period under consideration.

In their 2009 study, Van der Berg et al. used the same data and poverty line, and a similar adjustment on the income data was conducted, with the addition that there were small improvements in the techniques to estimate the distribution of wage income. The results of their analyses once again showed that the poverty headcount ratio increased between 1993 and 2000, before a continuous downward trend was observed between 2000 and 2004. With regard to the inequality trends, the Gini coefficient increased slightly during the period under investigation. Finally, the decomposition of the Theil-L and Theil-T indices by race showed that the contribution of the within-race component to the total increased continuously.

To conclude, these four studies using the AMPS data found that the poverty headcount ratio increased since 1993, but showed a continuous downward trend since 2000. In other words, the poverty trend was very similar to what was found in the censuses and IESs. However, contrary to the findings using other survey data, the Gini coefficient has always been hovering around the 0.68-0.70 range using AMPSs.

6.2.5 Review of studies that used at least two different survey data

In addition to the studies discussed above, other recent studies used at least two different sources of survey data to investigate poverty and inequality trends. First, Whiteford and Van Seventer (2000) used the Census 1996 data as well as the results of the earlier analyses of income distribution by McGrath (1983) and Whiteford and McGrath (1994) on data from 1975 and 1991 to investigate the changing income distribution of South Africa in the 1990s. They found that although the Gini coefficient increased slightly from 0.68 in 1991 to 0.69 in 1996, the contribution of within-race inequality to total inequality showed a continuous increase (38% in 1975, 58% in 1991 and 67% in 1996). The paper concluded that the prospects for a decline in inequality were not good, due to the increase in demand for well-paid highly-skilled people and decreased demand for lowly-paid less-skilled people who were forced into poorly remunerated informal sector employment or even into unemployment.

A UNDP (2003) study used the IES 1995 and a 2002 dataset[100] to derive the per capita expenditure for poverty and inequality analyses. The poverty headcount ratio increased from 0.09 to 0.11 using the $1 a day line, but increased if the other two poverty lines were used (from 0.24 to 0.23 and from 0.52 to 0.49 in the $2 a day and R354 per month in 1995 prices respectively). The Gini coefficient increased from 0.60 to 0.64 between the two surveys.

[100] The data source was not mentioned clearly by UNDP, but it was likely to be the LFS 2002b expenditure data.

Simkins (2004) used the IES 1995, IES 2000 as well as the Census 1996 and Census 2001 income data to investigate poverty and inequality trends. He applied a set of decision rules[101] to allocate positive incomes to some adults with unspecified incomes as well as to adults with zero personal incomes coming from households with zero household income (e.g., people who were age-eligible for social grants, ill/disabled people, employed, etc.). The imputed personal incomes were then used to derive household income. The per capita and per adult equivalent poverty lines that were consistent with a household income of R800 per month in 2000 prices were used[102]. The poverty headcount showed a clear increase between the two IESs and between the two censuses, regardless of whether per capita or per adult equivalent income variable was used. In addition, the Gini coefficient increased from just above 0.60 to nearly about 0.68.

The two studies by Meth (2006a and 2006b) focused on the validity of the claim by Van der Berg et al. (2005) that the number of poor decreased by 3.2 million between 2000 and 2004. Using various LFS and GHS expenditure data, he argued that the poverty headcount of 15.4 million in 2004 and even 18.6 million in 2000 as claimed by Van der Berg et al. were too low; he estimated that the decrease in number of poverty headcount between 2000 and 2004 was only approximately 1.5 million, using the R250 per month (2000 prices) poverty line. Both studies by Meth did not investigate inequality trends.

Leibbrandt, Woolard, Finn and Argent (2010) used the NIDS data in combination with the PSLSD and 2000 IES data to investigate poverty and inequality trends over time. However, certain income items were excluded (e.g., imputed rent, agricultural income, sale of vehicles and fixed property, etc.) so that the per capita income derived across the three data sources would be more comparable. Using such a revised per capita income variable, the analysis indicated that income inequality only worsened slightly during the period under investigation (the Gini coefficients being 0.66, 0.68 and 0.70 in 1993, 2000 and 2008 respectively). Income inequality was most serious amongst the black population residing in urban areas. In addition, the inequality decomposition showed that labour market income contributed a lot to income inequality (an approximately 80% contribution in all three data sources under study). As far as the poverty trends are concerned, using the proposed lower bound and upper bound poverty lines (R211 and R322 per month respectively), the poverty headcount ratio showed a negligent decrease – 0.56, 0.54 and 0.54 in 1993, 2000 and 2008 respectively using the R211

[101] The imputation method was, however, not explained in enough detail by Simkins in the paper.
[102] Simkins did not clearly specify these two poverty line income amounts in his study.

poverty line, and 0.72, 0.71 and 0.70 using the R322 poverty line.

Finally, the study by Posel and Rogan (2009) focused on poverty trends by gender. Using the data from OHS 1997, OHS 1999, GHS 2004 and GHS 2006, they derived two per capita income variables, with the total income being approximated as the sum of earned income and social grant income. For households having zero income as derived from this method, the per capita expenditure variable using the information from the household expenditure variable was used as proxy of per capita income. Using a poverty line of R322 per month in 2000 prices, the authors found that the poverty headcount ratio increased between 1997 and 1999, stagnated between 1999 and 2004, before showing a rapid decrease between 2004 and 2006. They argued that poverty remained a gendered phenomenon in post-apartheid South Africa, as the decline in poverty in the 2004-2006 period favoured males and male-headed households.

To conclude, the studies that used at least two different data sources found that the poverty headcount ratio increased in the 1990s, before showing a downward trend in the 2000s. The only exception is Leibbrandt et al. (2010), who found that the poverty headcount ratio showed a continuous but negligent decline since 1993. All studies found that, in general, the Gini coefficient increased moderately since the political transition.

Finally, in the study by Agüero, Carter and May (2005) that only focused on the KwaZulu-Natal province by using the expenditure information from the KwaZulu-Natal Income Dynamics Study (KIDS) data of 1993, 1998 and 2004, the poverty headcount ratio increased between 1993 and 1998 but decreased rapidly by 2004, while Gini coefficient increased continuously.

6.2.6 Summary

Table 6.1 summarizes the results of all these studies. Per capita variables were used in some studies while others used per adult equivalent variables; households with zero or unspecified income / expenditure were either excluded from the analyses or had the amounts imputed. The poverty lines used differed in the studies, but the three proposed official poverty lines were commonly used in recent studies. Despite these differences, the general conclusion is that the poverty headcount ratio increased moderately between 1994 and 2000, before a strong downward trend took place since 2000, in studies that used the censuses, IESs, AMPSs, or at least two different data sources. The only exception is that Leibbrandt et al. (2010) found that poverty showed a negligent decrease since the advent of democracy, when using the per capita income variables derived by excluding certain income items.

The Gini coefficient was found to increase between 1994 and 2000 rapidly in some studies but moderately in others. Since 2000, all studies found that the Gini coefficient still increased, but the extent of such increase was much less serious compared with what happened before 2000. Also, the Gini coefficients were higher in studies that used the census data but lower in studies that used the IES income or expenditure variables derived under the STC approach. Finally, with regard to within-race inequality and between-race inequality, the Theil-L or Theil-T indices showed that the within-race inequality as a share of total inequality increased continuously and became the dominant component since the transition according to all studies, except Bhorat and Van der Westhuizen (2008) and Bhorat et al. (2009).

Finally, for the studies discussed in Section 6.2, only a few of them dealt with the factors that could influence the comparability and reliability of the poverty and inequality estimates as discussed in Chapter 3. These issues will be taken into consideration when deriving poverty and inequality trends in Sections 6.3 and 6.4.

Table 6.1: Summary of the results of recent studies on poverty and inequality trends

Author(s)	**Data sources**	**Per capita variable**	**Poverty line(s)**	**Poverty headcount ratio**	**Gini coefficient**
			Census / CS 2007 data		
Ardington, Lam, Leibbrandt and Welch (2005)	Census 1996 & Census 2001 (before hot deck imputation)	Income (After SRMI2)	R124/month (2001 prices) R400/month (2001 prices)	1996: 0.38 2001: 0.42 1996: 0.60 2001: 0.68	1996: 0.72 2001: 0.82
Leibbrandt, Poswell, Naidoo and Welch (2006)	Census 1996 & Census 2001 (before hot deck imputation)	Income (Excluded zero-income households)	R250/month (1996 prices) US$2/day (1996 prices)	1996: 0.50 2001: 0.55 1996: 0.26 2001: 0.28	1996: 0.68 2001: 0.73
Leibbrandt, Poswell, Naidoo and Welch (2006)	Census 1996 & Census 2001 (before hot deck imputation)	Income (Included zero-income households)	R250/month (1996 prices) US$2/day (1996 prices)	1996: 0.59 2001: 0.65 1996: 0.40 2001: 0.44	1996: 0.74 2001: 0.79

Table 6.1: Continued

Author(s)	Data sources	Per capita variable	Poverty line(s)	Poverty headcount ratio	Gini coefficient
			IES data		
Stats SA (2002e)	IES 1995 & IES 2000	Income (STC)	N/A	N/A	1995: 0.56 2000: 0.57 (Household level)
Van der Berg and Louw (2004)	IES 1995 & IES 2000	Income (STC) (with adjustments)	R250/month (2000 prices)	1995: 0.39 2000: 0.38	N/A (Gini coefficients increased in all races)
Hoogeveen and Özler (2006)	IES 1995 & IES 2000	Expenditure (STC)	R174/month (2000 prices) R322/month (2000 prices)	1995: 0.32 2000: 0.34 1995: 0.58 2000: 0.58	1995: 0.57 2000: 0.58
Özler (2007)	IES 1995 & IES 2000	Expenditure (STC) (with adjustments)	R322/month (2000 prices)	1995: 0.58 2000: 0.58	1995: 0.57 2000: 0.58
Pauw and Mncube (2007)	IES 1995 & IES 2000	Income or expenditure (STC) (with adjustments)	US $1/day US $2/day R322/month (2000 prices)	1995: 0.05 2000: 0.11 1995: 0.22 2000: 0.31 1995: 0.45 2000: 0.52	1995: 0.62 2000: 0.66
Bhorat & Van der Westhuizen (2008)	IES 1995 & IES 2005	Consumption (COICOP)	R174/month (2000 prices) R322/month (2000 prices)	1995: 0.31 2005: 0.23 1995: 0.53 2005: 0.48	1995: 0.64 2005: 0.69
Bhorat, Van der Westhuizen and Jacobs (2009)	IES 1995 & IES 2005	Income (COICOP)	N/A	N/A	1995: 0.64 2005: 0.72
Yu (2008)	IES 1995, IES 2000 & IES 2005	Income (STC and COICOP approaches)	R322/month (2000 prices)	STC: 1995: 0.44 2000: 0.56 2005: 0.50 COICOP: 1995: 0.46 2000: 0.57 2005: 0.51	STC: 1995: 0.66 2000: 0.71 2005: 0.71 COICOP: 1995: 0.66 2000: 0.71 2005: 0.72
			OHS/LFS data		
Meth & Dias (2004)	OHS 1999 & LFS 2002b	Income (per adult equivalent)	R467 (per adult equivalent) (2000 prices)	N/A (Number of poor rose by 2 million)	N/A
Vermaak (2005)	OHS 1995, OHS 1997, LFS 2001b & LFS 2003b	Income from main job (per adult equivalent)	R152/month (per adult equivalent) (2000 prices) R405/month (per adult equivalent) (2000 prices)	1995: 0.37 1997: 0.49 2001: 0.51 2003: 0.53 1995: 0.55 1997: 0.63 2001: 0.66 2003: 0.67	1995: 0.61 1997: 0.67 2001: 0.62 2003: 0.63

Table 6.1: Continued

Author(s)	Data sources	Per capita variable	Poverty line(s)	Poverty headcount ratio	Gini coefficient
			AMPS data		
Van der Berg, Burger, Burger, Louw and Yu (2005)	AMPS 1993-2004	Income (with adjustments)	R250/month (2000 prices)	1993: 0.41 2000: 0.42 2004: 0.33	N/A (Gini coefficients increased for all races)
Van der Berg, Louw & Yu (2008)	AMPS 1993-2004	Income	R250/month (2000 prices)	1993: 0.50 1995: 0.52 2000: 0.51 2004: 0.47	1993: 0.67 1995: 0.67 2000: 0.68 2004: 0.68
Van der Berg, Louw & Du Toit (2008)	AMPS 1993-2006	Income	R250/month (2000 prices)	1993: 0.50 1995: 0.52 2000: 0.51 2004: 0.47 2006: 0.44	1993: 0.67 1995: 0.67 2000: 0.68 2004: 0.68 2006: 0.69
Van der Berg, Burger, Burger, Louw and Yu (2009)	AMPS 1993-2004	Income (with adjustments)	R250/month (2000 prices)	1993: 0.34 1995: 0.33 2000: 0.36 2004: 0.28	1993: 0.68 1995: 0.68 2000: 0.72 2004: 0.70
			Using various data sources		
Whiteford & Van Seventer (2000)	Census 1996 and the results from McGrath (1983) and Whiteford & McGrath (1994)	Income	N/A	N/A	1975: 0.68 1991: 0.68 1996: 0.69
UNDP (2003)	IES 1995 & LFS 2002b	Expenditure	US $1/day US $2/day R354/month (1995 prices)	1995: 0.09 2002: 0.11 1995: 0.24 2002: 0.23 1995: 0.52 2002: 0.49	1995: 0.60 2002: 0.64
Simkins (2004)	IES 1995, IES 2000, Census 1996 & Census 2001	Income (with adjustments)	A per capita poverty line and per adult equivalent poverty line that were consistent with a household income of R800 per month (2000 prices) [Note: the author did not clearly specify these two poverty line income amounts]	IESs: (per capita) 1995: 0.35 2000: 0.37 IESs (OECD adult equivalent): 1995: 0.34 2000: 0.38 IESs (EU adult equivalent): 1995: 0.32 2000: 0.36 Censuses (per capita): 1996: 0.35 2001: 0.37	IESs (household income): 1995: 0.61 2000: 0.67 IESs (OECD adult equivalent): 1995: 0.64 2000: 0.68 IESs (EU adult equivalent): 1995: 0.63 2000: 0.68 Censuses (per capita): 1996: 0.66 2001: 0.69

Table 6.1: Continued

Author(s)	Data sources	Per capita variable	Per capita poverty line(s)	Poverty headcount ratio	Gini coefficient
Using various data sources					
Meth (2006a)	GHS 2003, GHS 2004, LFS 2003b & LFS 2004b	Expenditure (with adjustments)	R250/month (2000 prices)	N/A (There were about 18-20 million poor people in 2004)	N/A
Meth (2006b)	LFS 2001b & LFS 2004b	Expenditure (with adjustments)	R250/month (2000 prices)	N/A (Number of poor decreased from 19.5 to 18 million between 2001 and 2004)	N/A
Posel and Rogan (2009)	OHS 1997, OHS 1999, GHS 2004 and GHS 2006	Income (with adjustments)	R322/month (2000 prices)	Income (earnings only): 1997: 0.65 1999: 0.68 2004: 0.69 2006: 0.64 Income (earnings and social grants): 1997: 0.63 1999: 0.66 2004: 0.65 2006: 0.59 Income (using expenditure as income proxy for zero-income households): 1997: 0.59 1999: 0.64 2004: 0.62 2006: 0.56	N/A
Leibbrandt, Woolard, Finn and Argent (2010)	PSLSD, IES 2000 (STC) & NIDS	Income (with adjustments)	R322/month (2000 prices) R593/month (2000 prices)	1993: 0.56 2000: 0.54 2008: 0.54 1993: 0.72 2000: 0.71 2008: 0.70	1993: 0.66 2000: 0.68 2008: 0.70
Using provincial data sources					
Agüero, Carter and May (2005)	KIDS 1993[#], 1998 and 2004	Expenditure	R322/month (2000 prices)	1993: 0.52 1998: 0.57 2004: 0.47	1993: 0.49 1998: 0.50 2004: 0.55

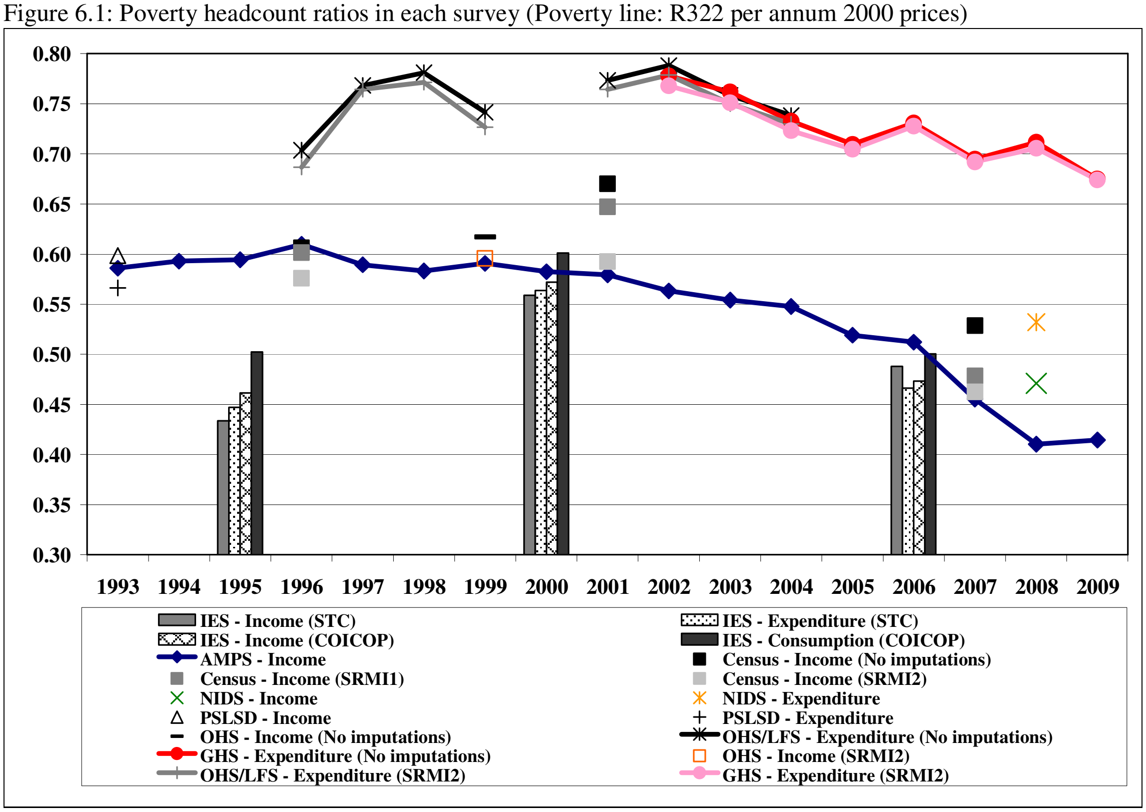
[#] The 1993 KIDS data was part of the PSLSD data.

6.3 Poverty trends since the transition

The post-SRMI per capita income and expenditure variables derived from Section 5.4 will now be used to investigate poverty and inequality trends. The FGT poverty measures in different surveys during the period under study are looked at, with the focus of the discussion being the poverty headcount ratios using the lower bound poverty line of R322 per month (2000 prices). The analysis takes place at person level, i.e., the product of household weight and household size is the weight variable.

Figure 6.1 and Table C.1 in Appendix C present the results in poverty headcount ratios between 1993 and 2009[103]. First, with regard to the poverty trends using the two censuses and CS 2007 but with no imputations applied on households with zero or unspecified household income, the poverty headcount ratio increased between 1996 and 2001, before a rapid decline took place between 2001 and 2007. The 2007 poverty headcount ratio was lower than the 1996 ratio.

Figure 6.1: Poverty headcount ratios in each survey (Poverty line: R322 per annum 2000 prices)



Poverty headcount ratios decreased in all three surveys after SRMI1, and such decrease was

103 The poverty headcount ratios by race at the R322 poverty line are shown in Table C.2 in Appendix C.

greater when SRMI2 was applied. However, the trends discussed above (i.e., upward trend between the two censuses, before a rapid downward trend took place between 2001 and 2007, with the 2007 poverty headcount ratio smaller than the 1996 result) remained the same, after SRMI1 or SRMI2 was conducted. It is possible that the extent of poverty increase between the two censuses was under-estimated because Census 1996 under-captured income more seriously than Census 2001 did, while the extent of the decline of poverty between 2001 and 2007 could be over-estimated because CS 2007 captured income much better (see Table 5.17 and Figure 5.12).

With regard to the poverty trends using the three IESs, the poverty headcount ratio increased rapidly between IES 1995 and IES 2000, before a downward trend was observed between IES 2000 and IES 2005/2006. This trend took place regardless of whether the STC or COCIOP approach was adopted. However, the IES 2005/2006 poverty headcount ratio was still slightly above the IES 1995 ratio. It was argued by Van der Berg, Louw and Du Toit (2008: 14) that the extent of increase of poverty could be over-estimated, since there was a large drop of recorded income (or expenditure) between IES 1995 and IES 2000 (while the national accounts income data showed that national income increased between the two years). Such a great drop in income between the two surveys was unlikely, as it was larger than the decrease experienced by the South African economy during the Great Depression of the 1930s. In addition, this decrease was also larger than the decline experienced by some of the affected countries during the 1998 Asian economic crisis. Thus, it seems certain issues (e.g., differences in sampling methodology) made the comparability of IES 1995 and IES 2000 difficult[104], and the poverty and inequality results between the two surveys should be interpreted with caution. Finally, since income (or expenditure) was very poorly captured in IES 2000, while IES 2005-2006 was the survey that captured income best, the extent of the decline of poverty between these two surveys could be over-estimated. It is possible that the use of the diary method could lead to better capturing of income and expenditure, and subsequently a lower poverty headcount ratio in IES 2005/2006.

Using the OHS and LFS per capita expenditure variable, the poverty headcount ratio increased since 1996, before a downward trend was observed from 2002. In addition, the 2004 poverty headcount ratio was slightly higher than the 1996 ratio. In the GHSs, a continuous downward trend in poverty was observed between 2002 and 2005, before an

[104] Two years after the publishing its report contrasting the results of IES 1995 and IES 2000, Stats SA admitted that the two surveys were not directly comparable.

unstable downward trend was observed between 2005 and 2009. The LFS 2002-2004 poverty headcount ratios were extremely close to the GHS 2002-2004 results. Furthermore, due to the serious under-estimation of total expenditure compared with national accounts as discussed in Section 5.5, the poverty headcount ratios in OHSs, LFSs and GHSs were much higher (always above 0.70) than the results using censuses and IESs (and also AMPS, NIDS and PSLSD, to be discussed below). It is also possible that the small number of expenditure categories in the OHSs, LFSs and GHSs partly contributed to the very high poverty estimates. After SRMI2 was applied in the OHS/LFS/GHS data, the same poverty headcount ratio trends as discussed above are still observed. However, the ratio shows a negligible decrease in the OHSs and LFSs, and the similar decrease is observed in the GHSs.

In AMPSs, it can be seen from Figure 6.1 that there was not too much change in the poverty headcount ratio before 2000, as it stabilised at approximately 0.59 between 1993 and 1999, before a continuous downward trend took place between 2000 and 2008. This trend is very different from what was found when looking at the censuses, IESs and OHSs, as these surveys indicated that poverty increased since the transition, before a downward trend took place since 2000. Nonetheless, the AMPS poverty headcount ratios in 1996, 2001 and 2007 were very close to the post-SRMI2 poverty headcount ratios in Census 1996, Census 2001 and CS 2007 respectively. Finally, the AMPS poverty headcount ratios have always been lower than the ratios derived in OHSs, LFSs and GHSs[105].

The PSLSD poverty headcount ratios, regardless of whether the income or expenditure variable was used, were very close to the 1993 AMPS ratio. However, even though total income was greater than total expenditure (Table 5.17), the poverty headcount ratio was higher for the income variable (0.598, compared with 0.566 when using the expenditure variable). In contrast, in NIDS, which also estimated a greater total income than expenditure, the poverty headcount ratio was higher if expenditure was used (0.532) while the ratio using income (0.471) was closer to the ratio in AMPS 2008 (0.410). Furthermore, as discussed in Section 5.5 (see Figures 5.14), there was no indication that expenditure was captured better than income when in surveys that collected both information, and hence for these surveys, the poverty headcount ratio using the per capita expenditure variable was not necessarily lower.

[105] This could be due to the relatively high proportions of households in the lower expenditure categories in these surveys. For instance, in AMPSs, the proportion of households with monthly household income of less than R1 200 decreased from 58.4% in 1993 to 23.5% in 2009, but this proportion hovered around the 60%-70% range in the OHSs and LFSs when looking at the monthly household expenditure (See Tables 2.18 and 5.13 as well as Figure C.1 in Appendix C), while this proportion only dropped from about 70% in GHS 2002 to just below 50% in GHS 2009. As another example, Figure C.2 shows the proportion of households with monthly household income or expenditure below R800 in each survey, and once again this proportion is lower in the AMPSs.

In order to determine how sensitive the poverty trends are to the poverty line chosen, dominance testing is required. Figures C.3 to C.16 in Appendix C show that, in general, in all surveys, the same poverty trends as discussed above are observed at all poverty lines, with two exceptions. First, in the IESs, the 1995 poverty headcount ratio was slightly above the 2005/2006 ratio at the lower poverty lines, but the opposite happened at the higher poverty lines, as indicated by the crossing of the IES 1995 and IES 2005/2006 lines in Figures C.7 to C.9 in Appendix C. Secondly, in PSLSD, the poverty headcount ratio was clearly lower if the expenditure variable was used at the lower poverty lines compared with what happened if the income variable was used, but the two lines crossed at approximately R6 000 per capita (Figure C.14). From then onwards, the poverty headcount ratio was always slightly higher for the expenditure variable (e.g., in the R592 poverty line, the poverty headcount ratio was 0.745 using the income variable, but 0.750 using the expenditure variable).

Figure 6.2: Poverty gap indices in each survey (Poverty line: R322 per month 2000 prices)

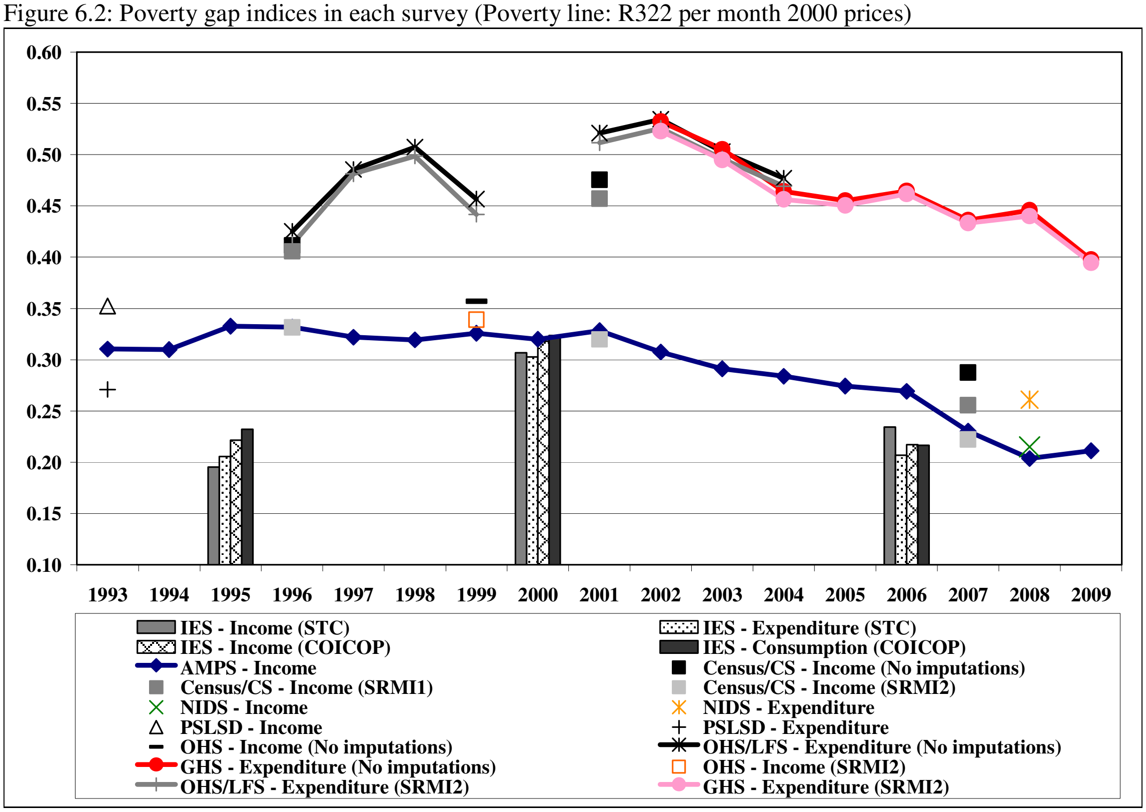


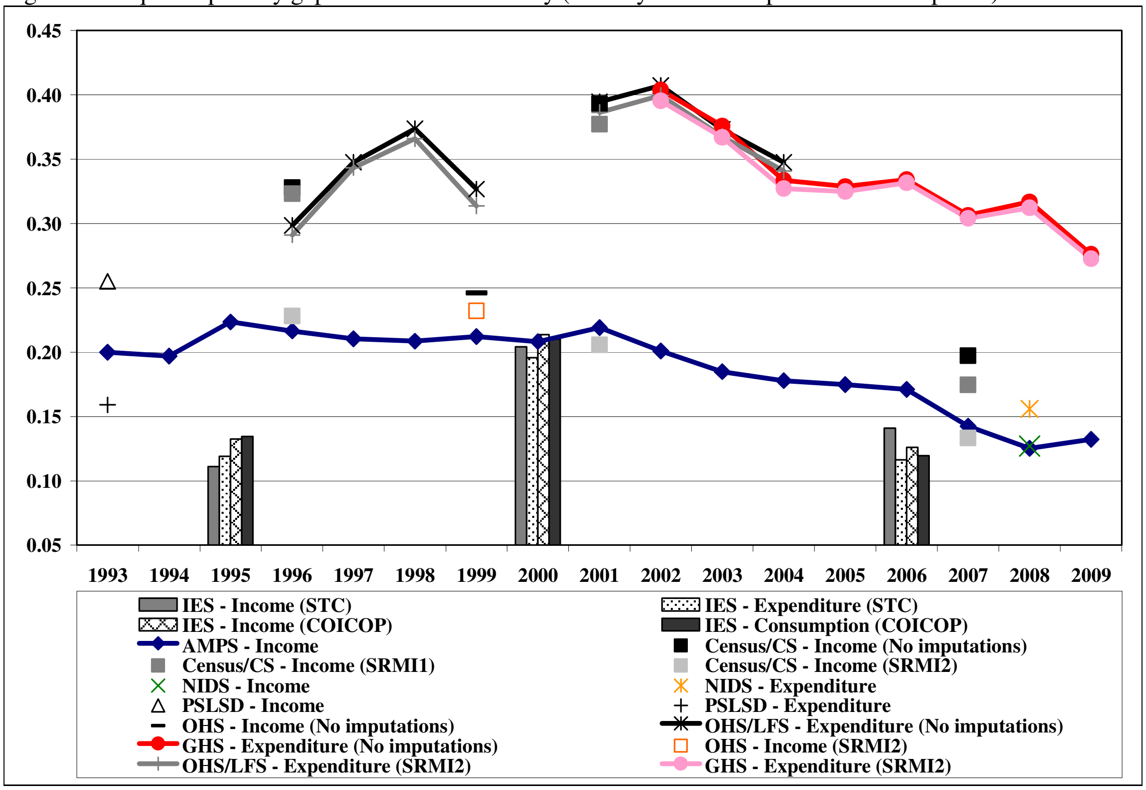
Figure 6.2 and Table C.3 in Appendix C present the results on poverty gap indices during the period under study, while Figure 6.3 and Table C.5 in Appendix C show the squared poverty gap indices[106]. With regard to the trends, very similar results as found when looking at

[106] Poverty gap indices and squared poverty gap indices by race at the R322 poverty line are presented in Tables C.4 and C.6 in Appendix C respectively.

poverty headcount ratios are observed. That is, all surveys under study found that poverty increased until about 2000, before a downward trend took place, with the exception of AMPS (which found that there was not much change in poverty between 1993 and 2001, before a downward trend took place between 2002 and 2008).

However, two findings are worth mentioning as far as the levels of these indices are concerned. First, in PSLSD, using the income and expenditure variables, the poverty headcount ratios were very close (0.598 and 0.566 respectively). However, there was quite a big difference when looking at the poverty gap indices (0.352 in the income variable and 0.271 in the expenditure variable) and squared poverty gap indices (0.255 vs. 0.159), that is, depth and severity of poverty in 1993 was more serious when using the income variable. Secondly, when comparing the poverty results of the Census 2001 income variable with no imputation, the Census 2001 income variable after SRMI1, and the LFS 2001 expenditure variable, the poverty headcount ratio using the latter variable was much larger than the results obtained when using the two census variables (see Figure 6.1). However, the poverty gap indices of the abovementioned three variables are closer (see Figure 6.2), and when looking at the squared poverty gap indices, it could be seen that these variables gave almost the same results (between 0.38 and 0.39 – see Figure 6.3).

Figure 6.3: Squared poverty gap indices in each survey (Poverty line: R322 per month 2000 prices)



Figures C.17 to C.44 in Appendix C show the sensitivity of the poverty trends, using poverty gap indices and squared poverty gap indices, to the poverty line chosen. The results show that, in general, in all surveys, the same poverty trends as discussed above are observed at all poverty lines, with few exceptions. First, after SRMI2 in the two censuses, the 1996 poverty gap index was always above the 2001 result at the lower poverty lines until about R4 800, from which the 2001 poverty gap index was higher (Figure C.19 in Appendix C). A similar result was found when comparing the squared poverty gap indices of the two post-SRMI2 census variables (Figure C.33 in Appendix C). At the poverty line of approximately R8000, the 2001 squared poverty gap index became greater.

Secondly, using the IES expenditure (STC) and income (COICOP) variables, the 1995 poverty gap and squared poverty gap indices were slightly higher than 2005/2006 ratios at the lower poverty lines, but the opposite took place at the higher poverty lines, as shown in Figures C.21, C.22, C.35 and C.36 in Appendix C. Finally, in PSLSD, it was found earlier that the poverty headcount ratio was lower if the expenditure variable was used at the lower poverty lines, but the ratio was higher using the income variable from R6 000 or above (Figure C.14), that is, poverty is subject to the poverty line chosen. However, with regard to poverty gap and squared poverty gap indices, Figures C.28 and C.42 clearly show that poverty was always higher for the income variable, regardless of the poverty line chosen.

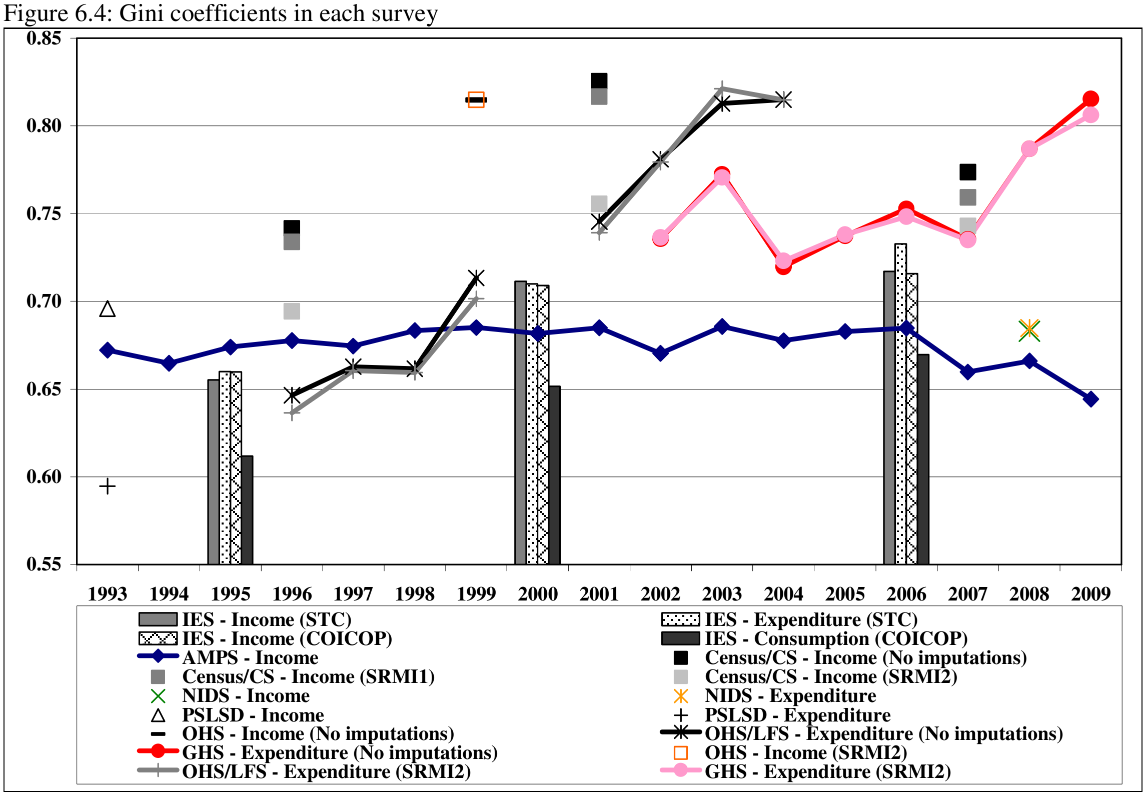
To conclude Section 6.3, despite the fact that the poverty levels differed amongst the surveys, using data from the censuses, CS 2007, IESs, OHSs, LFSs and GHSs, it was found that poverty increased since the transition, before a downward trend took place since 2001 or 2002. In other words, the findings were, in general, consistent with those in recent studies as discussed in Section 6.2. The only exception was that poverty level was stagnant between 1993 and 1999 if AMPS data was used. When looking at surveys that captured both income and expenditure information, there was no indication that poverty level was always lower using the per capita expenditure variable. That is, expenditure was not necessarily captured better and the subsequently poverty estimate was lower in a poorer, developing country, as discussed in Section 3.2. As the IES 2005/2006 was the only survey using the diary method (to complement the recall method), it is possible that this led to better capturing of income and expenditure information, and hence the decline of poverty between IES 2000 and IES 2005/2006. In addition, the findings in this section do not suggest that poverty levels were influenced by whether the income and expenditure was collected as continuous amounts or in interval terms. Furthermore, as expected, poverty levels decreased after the SMRI method was used to impute income of households reporting zero or unspecified amounts. Finally, for surveys that captured the income or expenditure information in fewer and wider intervals like OHSs/LFSs, GHSs, censuses and CS 2007, poverty levels were higher.

The impact of some of the factors discussed in Chapter 3 on the comparability and reliability of poverty levels and trends across the surveys will be examined in greater depth from Sections 6.5 onwards.

6.4 Inequality trends since the political transition

In this section, the Gini coefficients derived from the per capita variables are analysed. The results are presented in Figure 6.4[107] and Table 6.2. In the censuses, the Gini coefficient was measured to increase between Census 1996 and Census 2001, before a decline took place between Census 2001 and CS 2007. However, the 2007 value was slightly higher than the 1996 value. This pattern remained the same, regardless of whether imputation was conducted or not. However, the Gini coefficients in all three surveys became smaller after imputation was applied.

Figure 6.4: Gini coefficients in each survey



[107] The Gini coefficients by race of household head are presented in Table C.7 of Appendix C, while the Lorenz curves are shown in Figures C.45 to C.58.

Table 6.2: Gini coefficients in each survey

Survey	Per capita variable	Year	No imputations	After SRMI1	After SRMI2
Census/ CS	Income	1996	0.742	0.734	0.694
		2001	0.825	0.817	0.756
		2007	0.774	0.759	0.743
IES	Income – STC	1995	0.655	N/A	N/A
		2000	0.711		
		2005/2006	0.717		
	Expenditure – STC	1995	0.660		
		2000	0.710		
		2005/2006	0.733		
	Income - COICOP	1995	0.660	N/A	N/A
		2000	0.709		
		2005/2006	0.716		
	Consumption - COICOP	1995	0.612		
		2000	0.651		
		2005/2006	0.670		
OHS	Expenditure	1996	0.646	N/A	0.636
		1997	0.663		0.660
		1998	0.662		0.659
		1999	0.713		0.702
	Income	1999	0.815		0.815
LFS	Expenditure	2001	0.745	N/A	0.739
		2002	0.781		0.779
		2003	0.813		0.821
		2004	0.815		0.815
GHS	Expenditure	2002	0.736	N/A	0.736
		2003	0.772		0.771
		2004	0.720		0.723
		2005	0.737		0.738
		2006	0.753		0.748
		2007	0.735		0.735
		2008	0.787		0.787
		2009	0.815		0.806
PSLSD	Income	1993	0.696	N/A	N/A
	Expenditure	1993	0.595		
NIDS	Income	2008	0.683	N/A	N/A
	Expenditure	2008	0.685		
AMPS	Income	1993	0.672	N/A	N/A
		1994	0.665		
		1995	0.674		
		1996	0.678		
		1997	0.674		
		1998	0.683		
		1999	0.685		
		2000	0.682		
		2001	0.685		
		2002	0.670		
		2003	0.686		
		2004	0.678		
		2005	0.683		
		2006	0.685		
		2007	0.660		
		2008	0.666		
		2009	0.644		

A different trend was observed in the IESs, as there was a continuous worsening of income inequality, regardless of whether the STC or COICOP approach was adopted to derive the per capita variables. In addition, the increase of the Gini coefficient between IES 1995 and IES 2000 was more rapid, while there was only a slight increase in the coefficient between IES 2000 and IES 2005-2006. A similar trend was observed in the OHSs/LFSs, i.e., a continuous upward trend. However, the abrupt increase of Gini coefficient between the 2001 and 2003 LFSs (from just below 0.75 to about 0.82 – an increase of approximately 0.08) was peculiar. In the GHSs, an upward trend was also observed in general, although the Gini coefficient declined between 2003 and 2004, and again between 2006 and 2007. The extent of the very abrupt increase of the Gini coefficient between the 2007 and 2009 GHSs (from 0.74 to 0.82 – an increase of 0.08) was very similar to what was observed in the 2001-2003 LFSs as discussed above. In addition, there was no obvious change in the Gini coefficients as well as inequality trends, after SRMI2 was applied on both the OHS/LFS and GHS data.

In AMPSs, the Gini coefficients stayed within the 0.66-0.69 range between 1993 and 2007, before a slight downward trend took place in 2008-2009. This inequality trend was very different from what was found using other surveys. Furthermore, in the 1993 PSLSD, the Gini coefficient was much lower (0.595) if per capita expenditure was used but much higher (0.696) if per capita income was used. Finally, the Gini coefficient was similar (approximately 0.68) in the 2008 NIDS, regardless of which per capita variable was used. In addition, the 2008 NIDS Gini coefficient was quite close to the 2008 value using the AMPS per capita income variable.

When comparing the Gini coefficients across the different surveys, Table 6.2 shows that the Gini coefficients are more stable and lower in AMPSs, which involve more income categories and narrower bands (Table 2.23), while the estimates are greater in the two censuses, CS 2007, OHSs/LFSs, and GHSs, which involve fewer income / expenditure categories and wider bands (especially in the censuses and CS 2007). This seems to confirm the finding by Seiver (1979) that fewer and wider intervals are associated with greater measured inequality.

Looking at other indicators of inequality, Tables 6.3 and 6.4 as well as Figures C.59 to C.64 of Appendix C look at the racial decomposition of the Theil-L and Theil-T indices. The results show that the within-race inequality share of total inequality was more dominant and showed an obvious upward trend in the AMPSs. These findings are consistent with the results of the recent studies, as discussed in Section 6.2, with the exception of Bhorat and Van der

Westhuizen (2008) and Bhorat et al. (2009), who found the between-race inequality share to be more dominant.

Table 6.3: Theil-L indices in each survey

Survey	Per capita variable	Year	Theil-L index			Share: Within -race
			Within-race	Between-race	Total	
Census/ CS	Income – No imputations	1996	0.707	0.350	1.057	66.9%
		2001	0.893	0.395	1.288	69.3%
		2007	0.849	0.361	1.210	70.2%
	Income – After SRMI1	1996	0.682	0.344	1.026	66.5%
		2001	0.867	0.388	1.256	69.1%
		2007	0.810	0.352	1.162	69.7%
	Income – After SRMI2	1996	0.643	0.357	1.000	64.3%
		2001	0.744	0.428	1.172	63.5%
		2007	0.783	0.364	1.147	68.3%
IES	Income – STC	1995	0.520	0.311	0.831	62.6%
		2000	0.695	0.345	1.041	66.8%
		2005/2006	0.674	0.376	1.050	64.2%
	Expenditure – STC	1995	0.530	0.318	0.849	62.5%
		2000	0.637	0.393	1.030	61.9%
		2005/2006	0.687	0.401	1.088	63.1%
	Income - COICOP	1995	0.538	0.315	0.854	63.1%
		2000	0.686	0.340	1.026	66.8%
		2005/2006	0.631	0.405	1.037	60.9%
	Consumption - COICOP	1995	0.449	0.241	0.691	65.1%
		2000	0.503	0.314	0.818	61.5%
		2005/2006	0.491	0.354	0.845	58.1%
OHS	Expenditure – No imputations	1996	0.583	0.222	0.805	72.5%
		1997	0.482	0.353	0.836	57.7%
		1998	0.528	0.307	0.834	63.3%
		1999	0.626	0.369	0.995	62.9%
	Income – No imputations	1999	0.948	0.596	1.544	61.4%
	Expenditure – After SRMI2	1996	0.731	0.056	0.787	92.9%
		1997	0.471	0.359	0.830	56.8%
		1998	0.516	0.316	0.831	62.0%
		1999	0.596	0.377	0.974	61.2%
	Income – After SRMI2	1999	0.910	0.655	1.565	58.1%
LFS	Expenditure – No imputations	2001	0.689	0.449	1.138	60.6%
		2002	0.830	0.465	1.295	64.1%
		2003	0.844	0.664	1.509	56.0%
		2004	0.884	0.602	1.486	59.5%
	Expenditure – After SRMI2	2001	0.675	0.454	1.129	59.8%
		2002	0.815	0.486	1.302	62.6%
		2003	0.838	0.694	1.532	54.7%
		2004	0.869	0.621	1.490	58.3%
GHS	Expenditure – No imputations	2002	0.689	0.412	1.101	62.6%
		2003	0.840	0.420	1.259	66.7%
		2004	0.655	0.382	1.037	63.2%
		2005	0.710	0.401	1.111	63.9%
		2006	0.739	0.423	1.162	63.6%
		2007	0.743	0.349	1.092	68.0%
		2008	0.887	0.450	1.337	66.3%
		2009	1.047	0.491	1.539	68.1%

Table 6.3: Continued

Survey	Per capita variable	Year	Theil-L index			Share: Within -race
			Within-race	Between-race	Total	
GHS	Expenditure – After SRMI2	2002	0.679	0.430	1.109	61.2%
		2003	0.820	0.436	1.257	65.3%
		2004	0.649	0.406	1.055	61.6%
		2005	0.704	0.411	1.116	63.1%
		2006	0.727	0.419	1.145	63.4%
		2007	0.738	0.353	1.091	67.7%
		2008	0.878	0.464	1.342	65.4%
		2009	0.984	0.492	1.476	66.7%
PSLSD	Income	1993	0.553	0.455	1.008	54.8%
	Expenditure	1993	0.324	0.319	0.643	50.4%
NIDS	Income	2008	0.611	0.309	0.920	66.4%
	Expenditure	2008	0.591	0.330	0.922	64.2%
AMPS	Income	1993	0.444	0.461	0.905	49.1%
		1994	0.431	0.442	0.873	49.4%
		1995	0.484	0.442	0.927	52.3%
		1996	0.469	0.451	0.920	51.0%
		1997	0.484	0.430	0.914	53.0%
		1998	0.489	0.457	0.946	51.7%
		1999	0.503	0.447	0.951	53.0%
		2000	0.562	0.376	0.938	59.9%
		2001	0.576	0.388	0.964	59.8%
		2002	0.576	0.332	0.908	63.5%
		2003	0.573	0.371	0.944	60.7%
		2004	0.552	0.358	0.910	60.7%
		2005	0.620	0.326	0.946	65.5%
		2006	0.621	0.328	0.949	65.5%
		2007	0.614	0.267	0.881	69.7%
		2008	0.641	0.260	0.901	71.1%
		2009	0.671	0.190	0.861	78.0%

Table 6.4: Theil-T indices in each survey

Survey	Per capita variable	Year	Theil-L index			Share: Within -race
			Within-race	Between-race	Total	
Census/ CS	Income – No imputations	1996	0.594	0.427	1.021	58.2%
		2001	1.069	0.520	1.589	67.3%
		2007	0.992	0.488	1.480	67.0%
	Income – After SRMI1	1996	0.571	0.417	0.987	57.8%
		2001	1.021	0.493	1.514	67.4%
		2007	0.921	0.463	1.384	66.6%
	Income – After SRMI2	1996	0.565	0.442	1.006	56.1%
		2001	0.918	0.567	1.485	61.8%
		2007	0.891	0.479	1.369	65.1%
IES	Income – STC	1995	0.505	0.390	0.895	56.4%
		2000	0.640	0.454	1.094	58.5%
		2005/2006	0.637	0.505	1.141	55.8%
	Expenditure – STC	1995	0.509	0.400	0.908	56.0%
		2000	0.604	0.514	1.118	54.1%
		2005/2006	0.741	0.533	1.274	58.2%
	Income - COICOP	1995	0.517	0.395	0.913	56.7%
		2000	0.635	0.449	1.084	58.6%
		2005/2006	0.596	0.542	1.138	52.3%
	Consumption - COICOP	1995	0.498	0.302	0.800	62.2%
		2000	0.479	0.411	0.890	53.8%
		2005/2006	0.480	0.476	0.957	50.2%

Table 6.4: Continued

Survey	Per capita variable	Year	Theil-L index			Share: Within -race
			Within-race	Between-race	Total	
OHS	Expenditure – No imputations	1996	0.696	0.281	0.977	71.3%
		1997	0.649	0.456	1.105	58.7%
		1998	0.615	0.398	1.013	60.7%
		1999	0.670	0.489	1.158	57.8%
	Income – No imputations	1999	0.882	0.769	1.651	53.4%
	Expenditure – After SRMI2	1996	0.886	0.067	0.952	93.0%
		1997	0.624	0.458	1.081	57.7%
		1998	0.586	0.403	0.989	59.2%
		1999	0.628	0.486	1.114	56.4%
	Income – After SRMI2	1999	0.814	0.803	1.616	50.3%
LFS	Expenditure – No imputations	2001	0.714	0.592	1.306	54.7%
		2002	0.929	0.617	1.546	60.1%
		2003	0.958	0.869	1.826	52.4%
		2004	0.998	0.791	1.789	55.8%
	Expenditure – After SRMI2	2001	0.688	0.589	1.277	53.9%
		2002	0.893	0.630	1.523	58.6%
		2003	0.933	0.881	1.815	51.4%
		2004	0.962	0.803	1.764	54.5%
GHS	Expenditure – No imputations	2002	0.719	0.553	1.272	56.5%
		2003	0.928	0.573	1.500	61.8%
		2004	0.675	0.509	1.184	57.0%
		2005	0.760	0.534	1.295	58.7%
		2006	0.842	0.564	1.407	59.9%
		2007	0.815	0.460	1.275	63.9%
		2008	0.944	0.573	1.517	62.3%
		2009	0.987	0.635	1.622	60.8%
	Expenditure – After SRMI2	2002	0.703	0.565	1.268	55.4%
		2003	0.894	0.581	1.476	60.6%
		2004	0.660	0.530	1.190	55.5%
		2005	0.750	0.542	1.291	58.0%
		2006	0.821	0.554	1.375	59.7%
		2007	0.806	0.463	1.270	63.5%
		2008	0.924	0.582	1.506	61.3%
		2009	0.935	0.633	1.569	59.6%
PSLSD	Income	1993	0.485	0.549	1.034	46.9%
	Expenditure	1993	0.292	0.389	0.682	42.9%
NIDS	Income	2008	0.569	0.408	0.977	58.2%
	Expenditure	2008	0.545	0.426	0.972	56.1%
AMPS	Income	1993	0.350	0.549	0.899	38.9%
		1994	0.349	0.532	0.880	39.6%
		1995	0.366	0.533	0.899	40.8%
		1996	0.370	0.548	0.918	40.3%
		1997	0.382	0.524	0.906	42.2%
		1998	0.389	0.548	0.937	41.5%
		1999	0.400	0.538	0.938	42.7%
		2000	0.464	0.471	0.935	49.6%
		2001	0.456	0.480	0.937	48.7%
		2002	0.502	0.418	0.921	54.6%
		2003	0.550	0.463	1.013	54.3%
		2004	0.514	0.453	0.967	53.2%
		2005	0.597	0.415	1.012	59.0%
		2006	0.613	0.416	1.030	59.6%
		2007	0.574	0.342	0.917	62.6%
		2008	0.618	0.331	0.950	65.1%
		2009	0.555	0.240	0.795	69.8%

Figures C.65 to C.76 present the growth incidence curves (GICs) of IESs, censuses and AMPSs. First, comparing the IES 1995 and IES 2000 income variables (regardless of whether STC or COICOP approach is adopted), the GICs clearly showed that income growth was not pro-poor in both absolute sense and relative sense, as the curves (Figures C.65 and C.69) are below the horizontal axis and upward-sloping. That is, income decreased at all percentiles, but the decrease was more rapid at the poorer percentiles. The same finding was observed when comparing the 1995 and 2000 STC expenditure variables (Figure C.67) and COICOP consumption variables (Figure C.71). The abovementioned findings might explain why the measured poverty increased (See Figures 6.1-6.3) and inequality worsened between IES 1995 and IES 2000 (See Figure 6.4).

Comparing IES 2000 and IES 2005/2006, Figures C.66 and C.70 indicate that there was some pro-poor income growth in absolute sense took place, as the GICs are above the horizontal axis. In addition, the GICs are very steep and downward-sloping in the poorest 20% of the population, which imply that the income growth of the poor was greater. However, the income growth rates at the remaining 80 percentiles were very similar (as indicated by the very flat GICs). That might explain why the Gini coefficient was quite stable between the two surveys. Looking at the STC expenditure (Figure C.68) and COICOP consumption (Figure C.72) growth between the two surveys, the growth rate was greater in the lower percentiles (i.e., pro-poor growth in relative terms), but from the 80th percentile onwards, growth rate showed an upward trend (i.e., pro-poor growth in relative terms did not take place). This might explain why poverty decreased but inequality slightly worsened between IES 2000 and IES 2005/2006.

Looking at the post-SRMI2 income variables in the two censuses, Figure C.73 shows that there is an indication of pro-poor growth in absolute terms only in the poorest 40% and richest 10% of the population. In fact, negative income growth was experienced between the 40th and about 90th percentiles. In addition, the GIC was downward-sloping when looking at the poorest 60% of the population, but very steep and upward-sloping in the richest 10%. This indicates that pro-poor growth in relative terms only took place in the poorest 60%. This might account for the increase of both poverty and inequality between 1996 and 2001.

Comparing Census 2001 and CS 2007, Figures C.74 shows that income growth in absolute terms took place in all percentiles, but the growth was clearly slower in the poorest 20% of the population. The GIC was horizontal (despite the fluctuations) beyond the 20th percentile, with income growth hovering around the 30%-55% range. These findings might explain why poverty decreased but inequality only slightly worsened between 2001 and 2007.

Finally, with regard to the AMPS data, pro-poor growth in absolute sense took place between 1994 and 2004 (Figure C.75) and between 2004 and 2009 (Figure C.76), as the GICs are above zero. However, pro-poor growth in relative sense only took place significantly when comparing 2004 and 2009 (in the poorest 60% of the population).

To conclude Section 6.4, a continuous worsening of income inequality was observed in IESs, OHSs, LFSs and GHSs. Using the two censuses and the CS 2007, the Gini coefficient increased between 1996 and 2001, before it declined between 2001 and 2007. In contrast, in AMPS surveys, the Gini coefficient was extremely stable between 1993 and 2007 before a slight decline took place in 2008 and 2009. Thus, inequality trends differed a lot amongst the surveys. Furthermore, the Gini coefficients were relatively greater in the OHSs, LFSs and GHSs (which used fewer and narrower intervals) as well as the two censuses and CS 2007 (with very wide intervals in the high-income categories). Also, after adopting the SRMI method to deal with households with zero or unspecified income in the surveys, it was only in the two censuses and CS 2007 that the inequality levels decreased (i.e., SRMI had negligent impact on the inequality estimates in the OHSs, LFSs and GHSs), but the trends over time remained the same.

6.5 Dealing with the imputed rent variable

IES 2005 (COICOP approach) and NIDS are the only two surveys that took imputed or implied rent into consideration when deriving total income from the sum of the income from different sources. Table 6.5 shows imputed rent as proportion of income or consumption.

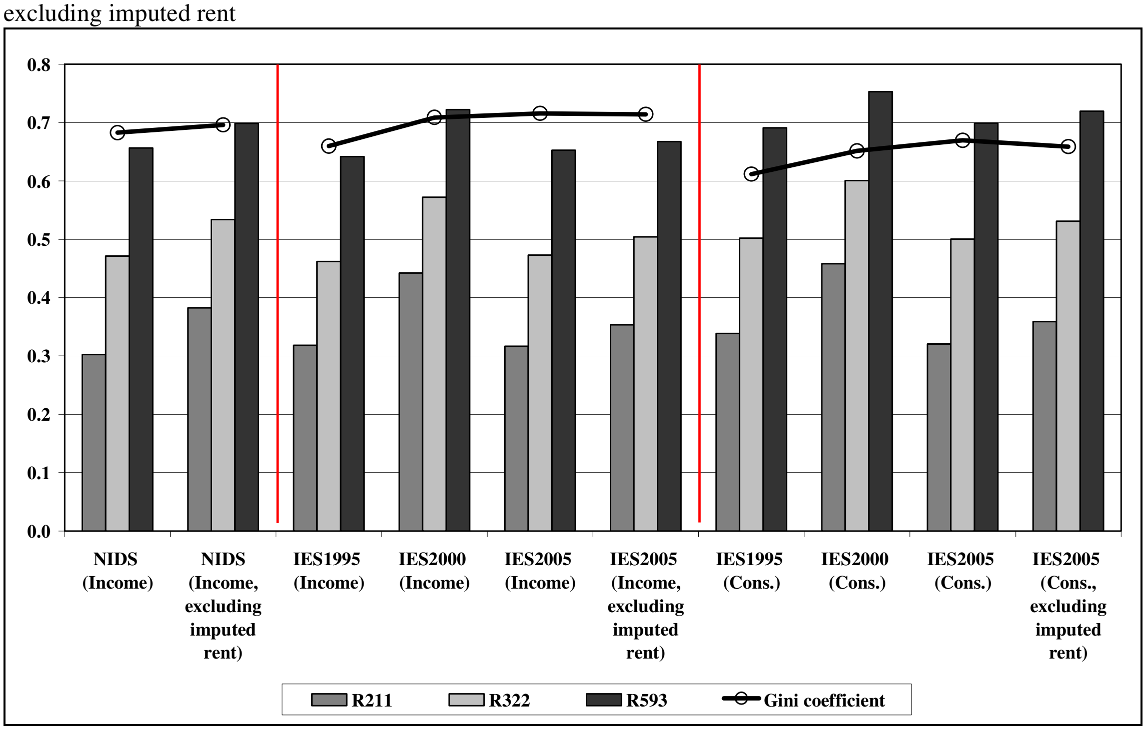
Table 6.5: Imputed rent amounts (Rand million, 2000 prices) and as percentage of total income, IES 2005/2006 and NIDS

	Imputed rent	Total income or consumption (Excluding imputed rent)	Total Income or consumption	Imputed rent as % of total income or consumption
NIDS	92 685	535 130	627 815	14.8%
IES 2005/2006 (Income)	66 927	638 786	705 713	9.5%
IES 2005/2006 (Consumption)	66 927	464 459	531 386	12.6%

As discussed in Section 6.2.5, the study by Leibbrandt et al. (2010) excluded certain items (with imputed rent being one of them) when deriving a revised per capita variable in PSLSD, IES 2000 and NIDS, and found that poverty showed a continuous but slight downward trend between 1993 and 2008, contradicting the findings in many other studies (upward trend in the 1990s before a decline took place since 2000-2002). Hence, this section briefly looks at whether the exclusion of the imputed rent variable significant changed the poverty and inequality estimates in IES 2005/2006 and NIDS.

The results are presented in Figure 6.5. It can be seen that the imputed rent variable has a negligible impact on Gini coefficients, but the coefficient increased slightly in NIDS after excluding imputed rent. The Gini coefficient also decreased slightly in IES 2005 after the exclusion of imputed rent. Moreover, as expected, after excluding imputed rents, the poverty headcount increased in both surveys. However, in the IESs, the same trend is still observed, i.e., an increase of the poverty headcount between 1995 and 2000, followed by a downward trend between 2000 and 2005. However, the 2005 headcount ratios (after excluding imputed rents) are now slightly higher than the 1995 ratios. Finally, with regard to the poverty trends, even after the exclusion of the imputed rent variable in IES 2005/2006, the same poverty trends were still observed, i.e., an increase between 1995 and 2000, before a decline took place in the 2000 and 2005/2006 IESs. That is, the poverty trend as found by Leibbrandt et al. (2010) did not take place in the three IESs after excluding imputed rent.

Figure 6.5: Poverty headcount ratios and Gini coefficients in IES 2005/2006 and NIDS, before and after excluding imputed rent

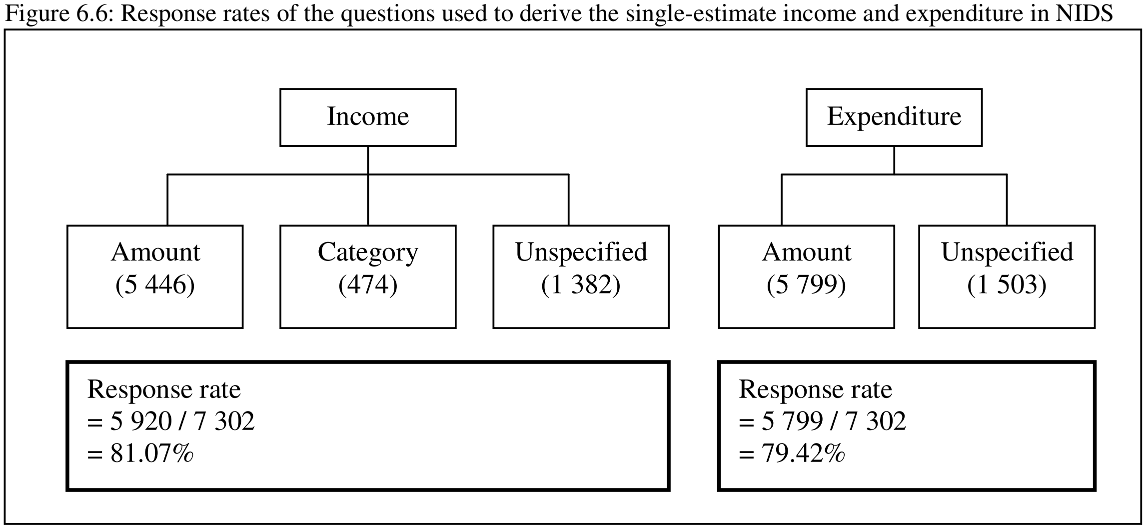


6.6 Comparing the single estimate and aggregated amount in NIDS

As discussed in Section 2.7.2.3, both the single estimate and aggregation approaches were adopted in NIDS to derive household income and expenditure. However, due to the relatively low response rate of the questions used in the single estimate approach (see Figure 6.6), and because SALDRU was concerned that income and expenditure could be under-captured in this approach, the income and expenditure derived from the aggregation approach were used as the variables for deriving poverty and inequality estimates. This section aims to investigate these estimates using the variables derived from both approaches, in order to find out whether the poverty and inequality levels and trends are significantly influenced by whether the single estimate, 'one-shot' method or aggregation method was used to capture the income and expenditure information (See Section 3.5).

Only households that provided responses to the questions relating to both single estimate and aggregation approaches (i.e., 5 920[108] households in the case of income and 5 799 households in the case of expenditure) are included for the analyses in Section 6.6, unless stated otherwise.

Figure 6.6: Response rates of the questions used to derive the single-estimate income and expenditure in NIDS



First, Figure 6.7 presents the total household income and expenditure derived from both methods. The single estimate approach resulted in lower recorded income and expenditure. In fact, the expenditure for the total population derived from this method (R237 364 million) is a

[108] In the case of income in the single estimate approach, the respondents were given the option to report the exact amount or the relevant income. With regard to the latter, the midpoint-Pareto method was used to make the data continuous – refer to Section 2.7.2.3.

staggering 49.14% lower than the amount derived from the aggregation method (R466 683 million). The lower amount captured in the single estimate approach might be attributed to the fact that the respondents did not know which items to include in this 'one-shot' amount (see the discussion in Section 3.5). On the other hand, the total income derived from the single estimate method (R429 561 million) was 19.18% lower than the amount derived from the aggregation approach (R531 525 million).

Figure 6.7: Total household income and expenditure (Rand million, 2000 prices) derived from single estimate and aggregation methods in NIDS

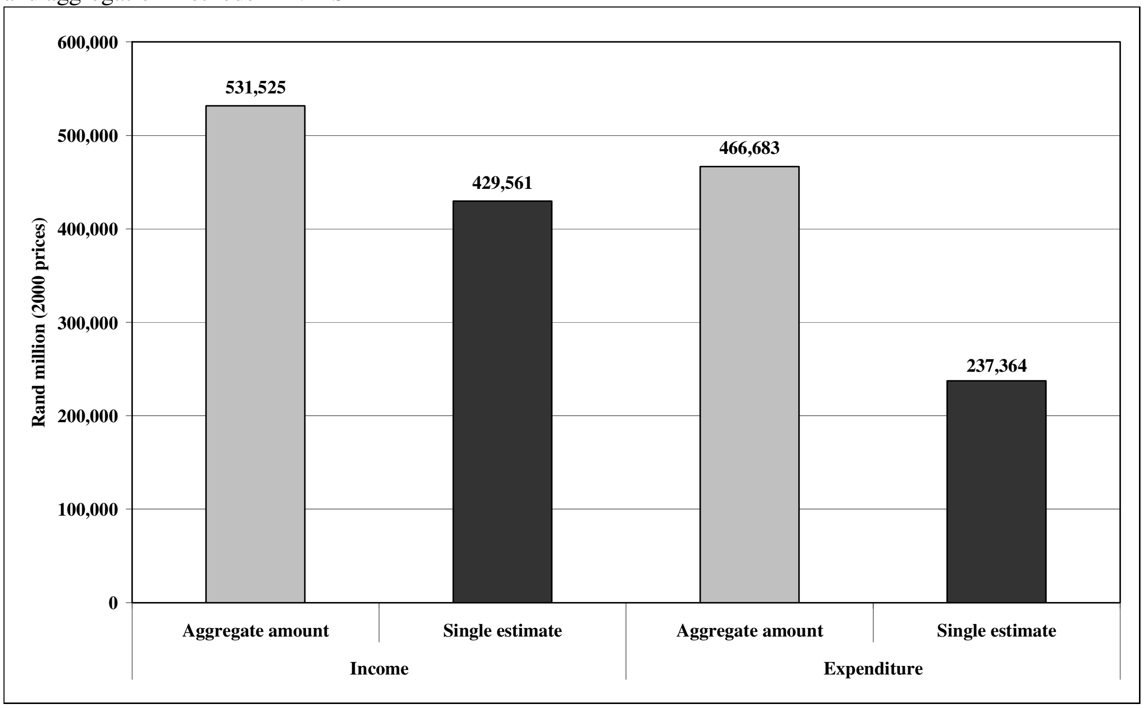


Figure 6.7 is complemented by Figures D.1 and D.2 in Appendix D which show the kernel density functions of log per capita income and expenditure from the two approaches. The results once again show that the single estimation approach seems to have under-estimated income and expenditure.

The 5 920 households reporting both the 'one-shot' and aggregate amounts in income are divided into deciles using the aggregate per capita income variable, and Table 6.6 shows that 18.8% of the households actually reported higher income in the single estimate (this proportion is the highest decile 1, which represents the households with the lowest per capita income). Moreover, 75.7% of households reported higher aggregate income amounts when compared with their 'one-shot' amounts.

Table 6.6: Comparison of single estimate and aggregate income amounts by decile, NIDS

		[1]	[2]	[3]	[4]	[5]	[6]	[7]		[3]-[7]
Household decile	Decile1	28.9%	7.0%	12.8%	21.6%	13.8%	5.0%	11.0%	100.0%	64.2%
	Decile2	20.1%	3.9%	21.0%	21.8%	13.6%	9.9%	9.8%	100.0%	76.0%
	Decile3	15.5%	3.1%	16.3%	24.9%	16.7%	12.8%	10.7%	100.0%	81.4%
	Decile4	13.6%	5.0%	21.9%	19.9%	16.3%	14.3%	9.1%	100.0%	81.5%
	Decile5	15.8%	2.9%	22.4%	17.0%	17.7%	10.7%	13.5%	100.0%	81.3%
	Decile6	15.6%	4.1%	30.8%	18.4%	13.0%	10.6%	7.6%	100.0%	80.3%
	Decile7	16.3%	7.4%	23.7%	22.2%	11.3%	11.8%	7.3%	100.0%	76.3%
	Decile8	20.4%	10.3%	24.5%	19.5%	12.8%	6.7%	5.8%	100.0%	69.3%
	Decile9	21.9%	5.9%	25.0%	18.8%	15.2%	5.8%	7.6%	100.0%	72.3%
	Decile10	19.9%	5.5%	20.6%	25.0%	13.3%	9.4%	6.3%	100.0%	74.7%
All households		18.8%	5.5%	21.9%	20.9%	14.4%	9.7%	8.9%	100.0%	75.7%

Note: household deciles are derived by using the aggregate per capita income variable.
[1]: Single estimate > Aggregate amount [2]: Single estimate = Aggregate amount
[3]: Single estimate < Aggregate amount: 0%-20% [4]: Single estimate < Aggregate amount: 20%-40%
[5]: Single estimate < Aggregate amount: 40%-60% [6]: Single estimate < Aggregate amount: 60%-80%
[7]: Single estimate < Aggregate amount: 80%-100%

Table 6.7 compares the 'one-shot' and aggregate expenditure amounts of the 5 799 households that reported both amounts (once again, by dividing households into deciles), and the results once again show that a relatively higher proportion of households from the poorer deciles had the 'one-shot' amount greater than the aggregate amount (e.g., 23.2% in decile1 and 17.8% in decile2). In addition, 87.7% of households reported higher aggregate expenditure amounts.

Table 6.7: Comparison of single estimate and aggregate expenditure amounts by decile, NIDS

		[1]	[2]	[3]	[4]	[5]	[6]	[7]		[3]-[7]
Household decile	Decile1	23.2%	0.1%	10.4%	20.1%	25.5%	17.9%	2.9%	100.0%	76.7%
	Decile2	17.8%	0.0%	10.2%	17.7%	23.4%	23.3%	7.7%	100.0%	82.2%
	Decile3	17.0%	0.0%	8.4%	15.3%	24.1%	28.4%	6.7%	100.0%	83.0%
	Decile4	10.5%	0.0%	8.8%	15.1%	24.3%	29.1%	12.2%	100.0%	89.5%
	Decile5	9.5%	0.0%	6.6%	15.9%	21.2%	36.6%	10.2%	100.0%	90.5%
	Decile6	9.7%	0.0%	9.0%	11.7%	22.0%	30.1%	17.5%	100.0%	90.3%
	Decile7	9.7%	0.0%	10.3%	15.1%	20.6%	27.5%	16.8%	100.0%	90.3%
	Decile8	10.6%	0.0%	4.9%	14.4%	23.9%	29.6%	16.6%	100.0%	89.4%
	Decile9	10.6%	0.0%	5.0%	20.0%	19.4%	21.6%	23.4%	100.0%	89.4%
	Decile10	4.9%	0.0%	9.6%	17.7%	24.7%	21.2%	21.9%	100.0%	95.1%
All households		12.4%	0.0%	8.3%	16.3%	22.9%	26.5%	13.6%	100.0%	87.7%

Note: household deciles are derived by using the aggregate per capita expenditure variable.
[1]: Single estimate > Aggregate amount [2]: Single estimate = Aggregate amount
[3]: Single estimate < Aggregate amount: 0%-20% [4]: Single estimate < Aggregate amount: 20%-40%
[5]: Single estimate < Aggregate amount: 40%-60% [6]: Single estimate < Aggregate amount: 60%-80%
[7]: Single estimate < Aggregate amount: 80%-100%

Table 6.8 presents the FGT poverty estimates and Gini coefficients. The single estimation approach is associated with higher poverty rates[109] and inequality, regardless of whether per capita income or expenditure variables are used. Looking at the poverty headcount ratio at the R322 poverty line, it increased from 0.212 (aggregation approach) to 0.363 (single estimate approach) using the income variable, and from 0.247 (aggregation approach) to 0.468 (single estimation approach) using expenditure. The Gini coefficient was 0.683 and 0.685 using the income and expenditure variables respectively, under the aggregation approach. However, the coefficient increased to 0.753 and 0.725 respectively for the single estimation method[110].

Table 6.8: FGT poverty estimates and Gini coefficients, NIDS: Single estimation approach vs. Aggregation approach

Poverty line: R211			
	P0	**P1**	**P2**
Income: Single estimate	0.493	0.259	0.171
Income: Aggregate amount	0.300	0.120	0.065
Expenditure: Single estimate	0.625	0.355	0.238
Expenditure: Aggregate	0.363	0.144	0.075
Poverty line: R322			
	P0	**P1**	**P2**
Income: Single estimate	0.614	0.363	0.255
Income: Aggregate amount	0.468	0.212	0.125
Expenditure: Single estimate	0.726	0.468	0.342
Expenditure: Aggregate	0.509	0.247	0.147
Poverty line: R593			
	P0	**P1**	**P2**
Income: Single estimate	0.728	0.505	0.392
Income: Aggregate amount	0.651	0.377	0.255
Expenditure: Single estimate	0.815	0.611	0.493
Expenditure: Aggregate	0.672	0.410	0.287
Gini coefficient			
Income: Single estimate	0.753		
Income: Aggregate amount	0.683		
Expenditure: Single estimate	0.725		
Expenditure: Aggregate	0.685		

In conclusion, the use of the single estimate per capita income and expenditure variables in NIDS is associated with higher poverty and inequality estimates. This implies that the income and expenditure information is captured more precisely, and the subsequently poverty and inequality estimates are more reliable if the aggregation method is used, despite the fact that a longer questionnaire and interview time are required, and fatigue might take place.

[109] Figures D.3 and D.4 in Appendix D present the cumulative density functions, and the results show that poverty headcount ratios are always higher using income and expenditure derived by the single estimation approach, regardless of the poverty lines chosen.
[110] Figures D.5 and D.6 in Appendix D once again show that inequality was serious when using the 'one-shot' variables, as the Lorenz curves of 'one-shot' income (Figure D.5) and 'one-shot' expenditure (Figure D.6) variables are further away from the line of equality, compared with the curves derived using the aggregated income and expenditure variables.

6.7 Impact of number and width of intervals on IES and NIDS poverty and inequality estimates

As mentioned in Chapter 2, the income / expenditure / consumption variables in the IES were reported as continuous variables (i.e., respondents declaring the exact amounts earned or spent), so one might ask if the poverty and inequality estimates and trends would differ, had the respondents been asked to report the relevant intervals instead. The debate on whether income (expenditure) should be captured as a continuous variable or by means of intervals was discussed in Section 3.4. In Section 3.6, it was found that the poverty and inequality estimates were very similar in the continuous and grouped datasets, providing the midpoint-Pareto method was used to make the data continuous (Von Fintel 2006; Malherbe 2007). With regard to the grouped income (expenditure) data, there was a lack of studies investigating the impact of the number and width of the intervals on poverty and inequality estimates, as discussed in Section 3.7.

Hence, the aim of this section is twofold: first, the AMPS 2000 intervals are applied on the actual continuous income variable (STC approach) of IES 2000, and the following methods are applied to make the dataset continuous again before poverty and inequality estimates are compared: (1) the midpoint method; (2) the midpoint method on all intervals, except the open interval, for which the Pareto method is applied; (3) the midpoint method on intervals up to and including the interval containing the population median income, and the Pareto method is used for the remaining intervals; (4) the interval regression method to derive the income of each household in intervals other than the zero income interval, by including the following explanatory variables: Province, area type, race of household head, gender of household head, age of household head, highest educational attainment of household head, employment status of household head, number of employed in the household, and household size; (5) the interval regression method, using the same explanatory variables as in (4) but adding two additional explanatory variables, namely the number of children aged 0-17 years and number of elderly aged at least 60 years in the household, as child grant and old-age grant are an important source of income at the lower end of the income distribution; (6) the interval regression method without any explanatory variables, that is, the regression is run on a constant only.

Method (2) is commonly applied in all recent South African literature. If this method gives similar poverty and inequality estimates as those obtained from using the actual continuous income variable, then it would suggest that the midpoint-Pareto method (if respondents were asked to declare the relevant interval) and the exact amount method yield comparable results.

The second aim of this section is to examine the impact of the number and width of the intervals on poverty and inequality estimates. First, some AMPS 2000 intervals are collapsed in order to investigate if these estimates would change significantly. Next, the Census 1996, Census 2001, AMPS 2000 and GHS 2009 intervals are applied in all three IESs to study the poverty and inequality trends across the three surveys. However, as mentioned in Section 3.4, if the income brackets are left unadjusted, a greater proportion of households would fall in the high-income categories in the more recent surveys due to the impact of inflation. Hence, the intervals above will be adjusted to 2000 prices before they are applied in all three IESs.

Table 6.9: FGT poverty estimates and Gini coefficients, after applying the AMPS income intervals on the IES 2000 income (STC approach) data, and using various methods to make the data continuous

		FGT poverty index			Gini coefficient
		P0	**P1**	**P2**	
Poverty line: R211 per month per annum (2000 prices)					
The actual continuous income variable		0.429	0.206	0.127	0.711
Method used to make the data continuous	Method (1)	0.422	0.202	0.123	0.681
	Method (2)	0.422	0.202	0.123	0.737
	Method (3)	0.423	0.202	0.123	0.738
	Method (4)	0.418	0.145	0.065	0.602
	Method (5)	0.405	0.148	0.007	0.603
	Method (6)	0.235	0.064	0.002	0.365
Poverty line: R322 per month per annum (2000 prices)					
The actual continuous income variable		0.559	0.307	0.204	Same as above
Method used to make the data continuous	Method (1)	0.562	0.303	0.200	
	Method (2)	0.562	0.303	0.200	
	Method (3)	0.562	0.304	0.201	
	Method (4)	0.588	0.271	0.151	
	Method (5)	0.578	0.269	0.153	
	Method (6)	0.590	0.192	0.084	
Poverty line: R593 per month per annum (2000 prices)					
The actual continuous income variable		0.710	0.462	0.342	Same as above
Method used to make the data continuous	Method (1)	0.713	0.458	0.339	
	Method (2)	0.713	0.458	0.339	
	Method (3)	0.713	0.460	0.339	
	Method (4)	0.754	0.462	0.317	
	Method (5)	0.754	0.458	0.316	
	Method (6)	0.861	0.450	0.268	

Table 6.9 above shows the FGT poverty estimates and Gini coefficients, after applying the AMPS 2000 intervals on the IES 2000 income data, and using various methods to make the dataset continuous again. In general, the first three methods yield comparable results as those obtained from using the actual continuous income variable captured in IES 2000 at all three poverty lines. However, the three interval regression methods result in lower poverty at the

R211 line but higher poverty in the R322 and R593 lines. With regard to the Gini coefficients, methods (1) and (4)-(6) under-estimated inequality (the extent of under-estimation was very serious in method (6)), while methods (2) and (3) slightly over-estimated inequality. Method (2) results in the coefficient that is closest to that obtained using the continuous income data. To conclude, the midpoint-Pareto yields the poverty and inequality estimates that are most comparable to the results obtained by using the underlying continuous data.

Having answered the first question (it is acceptable to apply the midpoint-Pareto method), the next issue has to do with whether the number and width of intervals have any influence on the poverty and inequality estimates. The AMPS intervals are collapsed as follows (Table 6.10):

- Method (1): Some of the low-income intervals are collapsed together so that these intervals become wider. There are 25 intervals in total after collapsing.
- Method (2): Some of the high-income intervals are collapsed together so that these intervals become wider. There are 21 intervals in total after collapsing.
- Method (3): Both the low-income and high-income intervals are collapsed. There are 11 intervals in total after collapsing, and this method results in the widest intervals.

Table 6.10: Collapsing selected AMPS 2000 intervals

Original intervals	Collapsed intervals (1)	Collapsed intervals (2)	Collapsed intervals (3)
R1-R199	R1-R199	R1-R199	R1-R499
R200-R299	R200-R399	R200-R299	
R300-R399		R300-R399	
R400-R499	R400-R599	R400-R499	
R500-R599		R500-R599	R500-R999
R600-R699	R600-R799	R600-R699	
R700-R799		R700-R799	
R800-R899	R800-R999	R800-R899	
R900-R999		R900-R999	
R1 000-R1 099	R1 000-R1 199	R1 000-R1 199	R1 000-R1 999
R1 100-R1 199			
R1 200-R1 399	R1 200-R1 399	R1 200-R1 399	
R1 400-R1 599	R1 400-R1 599	R1 400-R1 599	
R1 600-R1 999	R1 600-R1 999	R1 600-R1 999	
R2 000-R2 499	R2 000-R2 499	R2 000-R3 999	R2 000-R3 999
R2 500-R2 999	R2 500-R2 999		
R3 000-R3 999	R3 000-R3 999		
R4 000-R4 999	R4 000-R4 999	R4 000-R5 999	R4 000-R5 999
R5 000-R5 999	R5 000-R5 999		
R6 000-R6 999	R6 000-R6 999	R6 000-R7 999	R6 000-R7 999
R7 000-R7 999	R7 000-R7 999		
R8 000-R8 999	R8 000-R8 999	R8 000-R9 999	R8 000-R9 999
R9 000-R9 999	R9 000-R9 999		
R10 000-R10 999	R10 000-R10 999	R10 000-R11 999	R10 000-R11 999
R11 000-R11 999	R11 000-R11 999		
R12 000-R13 999	R12 000-R13 999	R12 000-R15 999	R12 000-R15 999
R14 000-R15 999	R14 000-R15 999		
R16 000-R17 999	R16 000-R17 999	R16 000-R19 999	R16 000-R19 999
R18 000-R19 999	R18 000-R19 999		
R20 000+	R20 000+	R20 000+	R20 000+

The FGT poverty indices and Gini coefficients are reported in Table 6.11. First, the poverty indices at all three poverty lines as well as the Gini coefficient by collapsing the low-income intervals (method 1) are almost the same as the results obtained by using the original AMPS intervals. However, the indices and the Gini coefficient only decreased after collapsing the high-income intervals (method 2) and declined slightly further after collapsing both low-income and high-income intervals (method 3). Thus, it seems poverty and inequality estimates only showed negligent changes after the application of fewer and wider intervals. That is, the results are contrary to what was found by Seiver (1979) in his study, i.e., the measured inequality increased by using fewer and wider intervals.

Table 6.11: FGT poverty estimates and Gini coefficients, after applying the AMPS income intervals on the IES 2000 income (STC approach) data, and collapsing some intervals

	FGT poverty index			**Gini coefficient**
	P0	**P1**	**P2**	
Poverty line: R211 per month per annum (2000 prices)				
The actual continuous income variable	0.429	0.206	0.127	0.711
AMPS intervals	0.422	0.202	0.123	0.737
AMPS collapsed intervals (1)	0.411	0.204	0.124	0.738
AMPS collapsed intervals (2)	0.414	0.201	0.123	0.731
AMPS collapsed intervals (3)	0.409	0.186	0.111	0.725
Poverty line: R322 per month per annum (2000 prices)				
The actual continuous income variable	0.559	0.307	0.204	Same as above
AMPS intervals	0.562	0.303	0.200	
AMPS collapsed intervals (1)	0.562	0.305	0.202	
AMPS collapsed intervals (2)	0.547	0.300	0.199	
AMPS collapsed intervals (3)	0.534	0.286	0.185	
Poverty line: R593 per month per annum (2000 prices)				
The actual continuous income variable	0.710	0.462	0.342	Same as above
AMPS intervals	0.713	0.458	0.339	
AMPS collapsed intervals (1)	0.713	0.459	0.340	
AMPS collapsed intervals (2)	0.713	0.452	0.334	
AMPS collapsed intervals (3)	0.703	0.443	0.323	

AMPS collapsed intervals (1): The low-income intervals are collapsed together
AMPS collapsed intervals (2): The high-income intervals are collapsed together
AMPS collapsed intervals (3): Both the low-income and high-income intervals are collapsed together

Next, in addition to the AMPS 2000 intervals, the following intervals are also applied on the IES 2000 income data, before investigating the poverty and inequality estimates: Census 1996 intervals, Census 2001 intervals, GHS 2009 intervals, equal R500 intervals, equal R1 000 intervals and equal R2 000 intervals. For the latter three approaches, the open interval is "R20 000+". However, due to the impact of inflation, the nominal intervals of the two censuses and GHS 2009 are converted to intervals in 2000 prices (See Table 6.12), before examining the poverty and inequality estimates obtained from the application of these intervals. Furthermore, the midpoint method was applied on all intervals, except that Pareto method was used for the open interval, in order to make each dataset continuous again.

Table 6.12: Adjusting the Census 1996, Census 2001 and GHS 2009 nominal intervals into 2000 prices intervals

Nominal intervals	**Real intervals**
Census 1996	
1: None	1: None
2: R1 – R200	2: R1 – R251
3: R201 – R500	3: R251 – R627
4: R501 – R1 000	4: R627 – R1 254
5: R1 001 – R1 500	5: R1 254 – R1 880
6: R1 501 – R2 500	6: R1 880 – R3 134
7: R2 501 – R3 500	7: R3 134 – R4 387
8: R3 501 – R4 500	8: R4 387 – R5 641
9: R4 501 – R6 000	9: R5 641 – R7 521
10: R6 001 – R8 000	10: R7 521 – R10 028
11: R8 001 – R11 000	11: R10 028 – R13 788
12: R11 001 – R16 000	12: R13 788 – R20 055
13: R16 001 – R30 000	13: R20 055 – R37 603
14: R30 001 or more	14: R37 603+
Census 2001	
1: None	1: None
2: R1 – R400	2: R1 – R377
3: R401 – R800	3: R377 – R754
4: R801 – R1 600	4: R754 – R1 509
5: R1 601 – R3 200	5: R1 509 – R3 017
6: R3 201 – R6 400	6: R3 017 – R6 035
7: R6 401 – R12 800	7: R6 035 – R12 070
8: R12 801 – R25 600	8: R12 070 – R24 140
9: R25 601 – R51 200	9: R24 140 – R48 279
10: R51 201 – R102 400	10: R48 279 – R96 558
11: R102 401 – R204 800	11: R96 588 – R193 117
12: R204 801 or more	12: R193 117+
GHS 2009	
1: R0	1: R0
2: R1 – R199	2: R1 – R115
3: R200 – R399	3: R115 – R231
4: R400 – R799	4: R231 – R461
5: R800 – R1 199	5: R461 – R692
6: R1 200 – R1 799	6: R692 – R1 038
7: R1 800 – R2 499	7: R1 038 – R1 442
8: R2 500 – R4 999	8: R1 442 – R2 884
9: R5 000 – R9 999	9: R2 884 – R5 767
10: R10 000 or more	10: R5 767+

Table 6.13 and Figure 6.8 report the poverty headcount ratios at the three poverty lines and Gini coefficients respectively by applying the different intervals to the STC income variable of the three IESs. With regard to the former, it can be seen that the FGT poverty indices at all three poverty lines when applying the AMPS 2000 and the R1 000 intervals are closest to those obtained by using the IES actual continuous income variable. Furthermore, poverty is clearly lower if the R2 000 intervals are applied. This could be explained by the fact that these R2 000 intervals are much wider at the lower end of the distribution (e.g., “R0 – R1 999”, “R2 000 – R3 999”, etc.). Hence, the income of the poor households could be over-estimated,

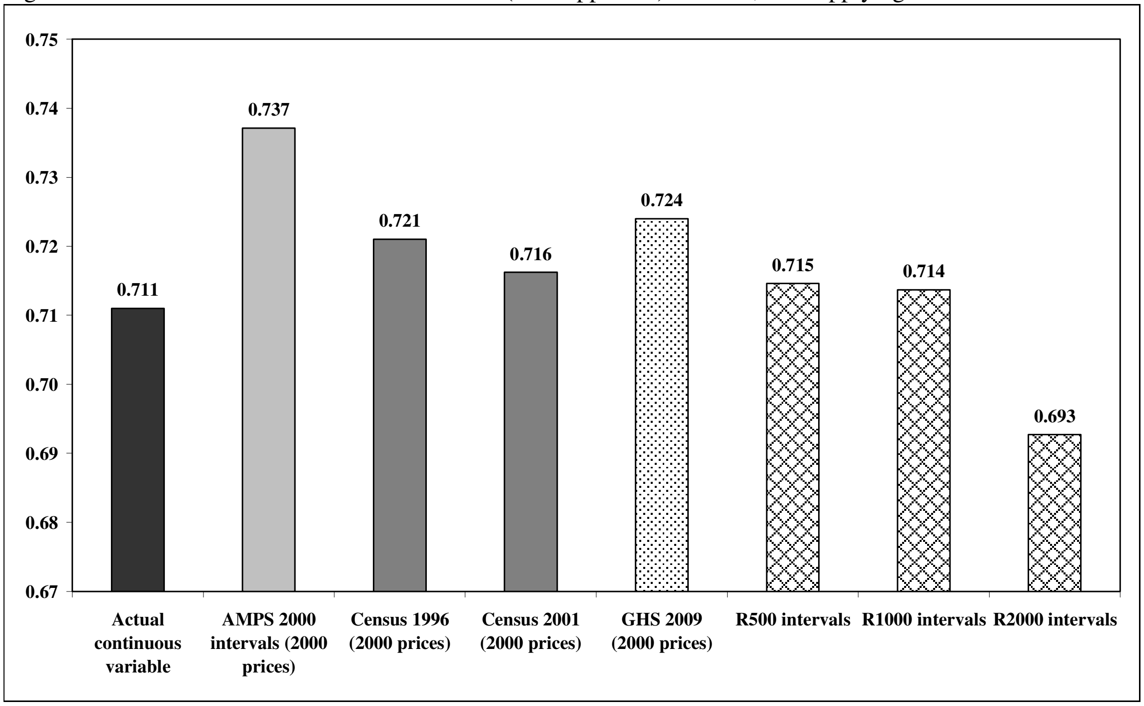
which eventually causes the under-estimation of poverty. Figures E.1 and E.2 in Appendix E provide more information by showing the CDFs of the income variables after applying various intervals, and it can be seen from Figure E.2 that the poverty headcount ratios are lower at the lower poverty lines, after applying the R2 000 intervals.

Table 6.13: FGT poverty indices, after applying various intervals on the IES 2000 income (STC approach) data

		FGT poverty index		
		P0	**P1**	**P2**
Poverty line: R211 per month per annum (2000 prices)				
The actual continuous income variable		0.429	0.206	0.127
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.422	0.202	0.123
	Census 1996 intervals (2000 prices)	0.417	0.207	0.129
	Census 2001 intervals (2000 prices)	0.412	0.198	0.123
	GHS 2009 intervals (2000 prices)	0.411	0.198	0.123
	R500 intervals	0.416	0.192	0.114
	R1 000 intervals	0.426	0.199	0.116
	R2 000 intervals	0.391	0.127	0.055
Poverty line: R322 per month per annum (2000 prices)				
The actual continuous income variable		0.559	0.307	0.204
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.562	0.303	0.200
	Census 1996 intervals (2000 prices)	0.569	0.306	0.205
	Census 2001 intervals (2000 prices)	0.538	0.297	0.197
	GHS 2009 intervals (2000 prices)	0.551	0.295	0.197
	R500 intervals	0.559	0.296	0.192
	R1 000 intervals	0.553	0.300	0.195
	R2 000 intervals	0.497	0.241	0.133
Poverty line: R593 per month per annum (2000 prices)				
The actual continuous income variable		0.710	0.462	0.342
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.713	0.458	0.339
	Census 1996 intervals (2000 prices)	0.705	0.459	0.340
	Census 2001 intervals (2000 prices)	0.717	0.448	0.331
	GHS 2009 intervals (2000 prices)	0.695	0.448	0.331
	R500 intervals	0.701	0.454	0.333
	R1 000 intervals	0.709	0.455	0.334
	R2 000 intervals	0.706	0.417	0.284

With regard to income distribution, Figure 6.8 shows that the Gini coefficients obtained by applying the Census 2001 intervals as well as the R500 and R1 000 intervals are in line with the Gini coefficient obtained by using the continuous dataset, hovering at the 0.71-0.72 range. The use of the wider R2 000 intervals resulted in a relatively serious under-estimation of inequality, while the opposite happened when using the AMPS 2000 intervals. The results do not conform to Seiver's (1979) findings that fewer and wider intervals are associated with greater inequality.

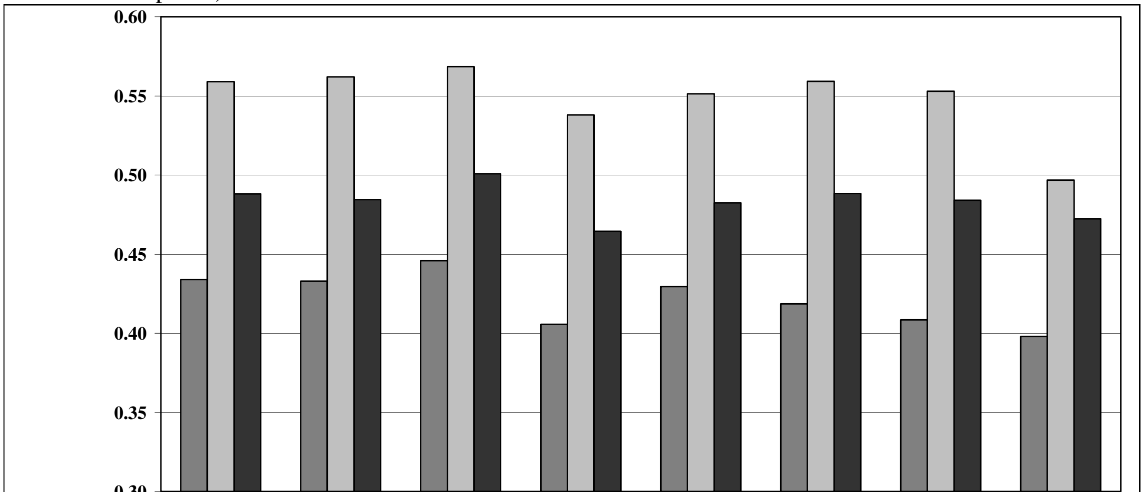
Figure 6.8: Gini coefficients on IES 2000 income (STC approach) variable, after applying various intervals



The intervals mentioned above are now also applied on the STC income variable of the other two IESs so as to investigate the poverty and inequality trends across the three IESs[111]. First, Figure 6.9 and Table E.2 in Appendix E present the FGT poverty indices across the three IESs. Focusing on the poverty headcount ratios at the R322 poverty line, it can be seen that the same poverty trends (i.e., rapid increase between IES 1995 and IES 2000, before a decline took place in IES 2005/2006, but the IES 2005/2006 poverty headcount ratio was still above the IES 1995 ratio) are observed, regardless of which intervals were applied. In addition, the poverty headcount ratios obtained by using the AMPS 2000, GHS 2009 and R500 intervals are closest to the results obtained by using the original continuous variable in all three surveys, but poverty was seriously under-estimated with the application of the Census 2001 and R2 000 intervals (fewer and wider intervals).

[111] Table E.1 in Appendix E shows the proportion of households in each income category in each IES, after the application of various intervals.

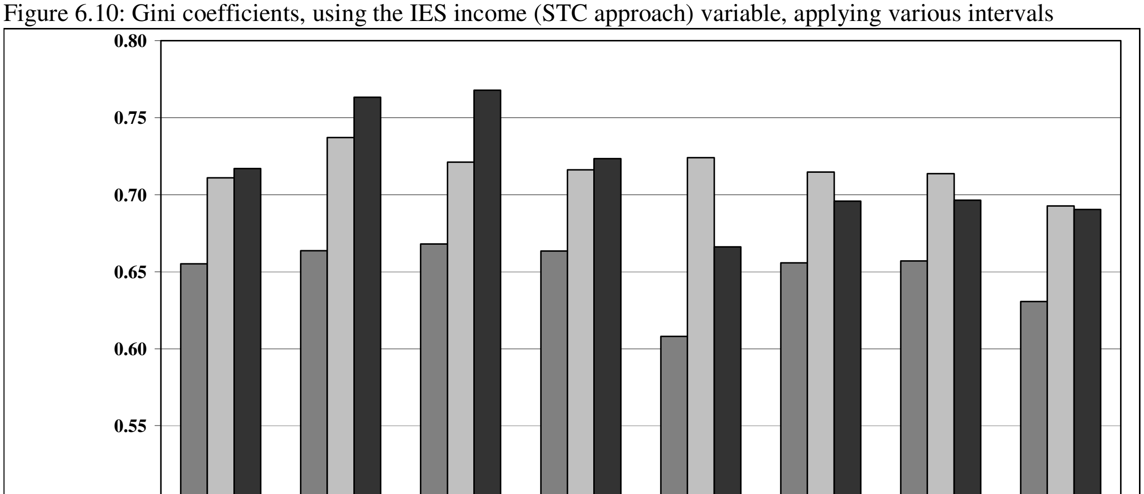
Figure 6.9: Poverty headcount ratios, after applying various intervals on the three IESs (Poverty line: R322 per month in 2000 prices)



	Actual continuous variable	AMPS 2000 intervals (2000 prices)	Census 1996 intervals (2000 prices)	Census 2001 intervals (2000 prices)	GHS 2009 intervals (2000 prices)	R500 intervals	R1000 intervals	R2000 intervals
IES1995	0.434	0.433	0.446	0.406	0.430	0.419	0.408	0.398
IES2000	0.559	0.562	0.569	0.538	0.551	0.559	0.553	0.497
IES2005/2006	0.488	0.484	0.501	0.464	0.482	0.488	0.484	0.472

The Gini coefficients across the three IESs are presented in Figure 6.10. Only the AMPS 2000, Census 1996 and Census 2001 intervals obtained inequality trends similar to the original continuous variable, i.e., a rapid increase between IES 1995 and IES 2000, followed by a smaller increase between IES 2000 and IES 2005/2006. The use of the GHS 2009, R500, R1 000 and R2 000 intervals resulted in a decrease of inequality between the last two IESs, and the decline was quite rapid with the use of the GHS 2009 intervals.

Figure 6.10: Gini coefficients, using the IES income (STC approach) variable, applying various intervals



	Actual continuous variable	AMPS 2000 intervals (2000 prices)	Census 1996 intervals (2000 prices)	Census 2001 intervals (2000 prices)	GHS 2009 intervals (2000 prices)	R500 intervals	R1000 intervals	R2000 intervals
IES1995	0.655	0.664	0.668	0.663	0.608	0.656	0.657	0.631
IES2000	0.711	0.737	0.721	0.716	0.724	0.715	0.714	0.693
IES2005/2006	0.717	0.763	0.768	0.723	0.666	0.696	0.696	0.690

A similar exercise was conducted on the NIDS single estimate as well as aggregated income and expenditure variables, as the AMPS 2007, CS 2007 and GHS 2007 intervals were applied, and then midpoint-Pareto method as discussed above was used to make the NIDS dataset continuous again. The results, as presented in Table E.3 as well as Figures E.3 – E.6 in Appendix E are very similar to what happened in the three IESs as investigated above, i.e., there were only slight to moderate changes to the estimates and trends of poverty headcount ratios: the use of AMPS and CS intervals resulted in very small decrease of poverty headcount ratios, while the use of the GHS bands (with very few intervals) led to a more moderate decline in the ratios (see Figures E.3 and E.5).

With regard to the inequality results, it is interesting that the use of AMPS intervals resulted in small reduction in the Gini coefficients, compared with what happened using the actual continuous variables as they are (Figures E.4 and E.6). The use of the GHS bands clearly resulted in quite a large increase of the coefficients, especially the increase from 0.725 (if the actual 'one-shot' expenditure was used) to 0.845 (after applying the GHS intervals). These results are consistent with what Seiver (1979) found, i.e., the use of fewer intervals lead to increased inequality estimates.

All the results discussed in Section 6.7 above should be interpreted with caution, as it is assumed that the respondents, who declared their income or expenditure by aggregation approach in the IESs, would report <u>similar</u> income or expenditure if asked to report the 'one-shot', categorical answer. For instance, if the aggregated income of a respondent in NIDS is equal to R4 400, then it is assumed that he/she would report that his income falls in the 'R4 000 – R4 999' interval, and then the midpoint method is applied by converting his/her categorical answer into an amount of R4 500 (which is quite close to the original actual amount of R4 400, and hence this would not have significant impact on poverty and inequality estimates). However, Sections 3.5 and 6.6 have shown that the respondents could give very different answers on their income or expenditure, when comparing the single estimate and aggregation approaches.

6.8 Poverty trends revisited after adjustments of survey distribution in line with national accounts income mean

This section will briefly investigate what would have happened to the poverty levels and trends, had the survey income / consumption / expenditure been adjusted in line with the national accounts income mean (i.e., shifting the distribution rightwards in all surveys except the post-SRMI2 income variable of OHS 1999, since these surveys all under-estimated income / consumption / expenditure, compared with the national accounts income of the same year, while OHS 1999 is the only one that over-estimated income). Note that inequality estimates will be unaffected as a result of the adjustments.

Table 6.14 and Figures F.1-F.6 in Appendix F report the findings. The results show the lower poverty headcount ratios after the adjustment, with the exception of the post-SRMI2 income variable of OHS 1999, which experienced a slight increase in the poverty headcount ratios at all three poverty lines after the adjustment. The poverty trends in each survey after the adjustment remain the same, as discussed in Section 6.3.

Table 6.14: Poverty headcount ratios at different poverty lines with and without adjustment of survey means in line with national accounts means, using the per capita variables

Survey	Per capita variable	Year	R211		R322		R593	
			[A]	[B]	[A]	[B]	[A]	[B]
Census/ CS	Income – No imputations	1996	0.501	0.345	0.606	0.445	0.728	0.589
		2001	0.568	0.406	0.670	0.518	0.789	0.668
		2007	0.397	0.270	0.529	0.409	0.694	0.610
	Income – After SRMI1	1996	0.493	0.242	0.601	0.326	0.726	0.462
		2001	0.547	0.335	0.647	0.432	0.769	0.568
		2007	0.351	0.219	0.478	0.335	0.656	0.559
	Income – After SRMI2	1996	0.441	0.257	0.576	0.334	0.715	0.481
		2001	0.447	0.336	0.592	0.440	0.750	0.592
		2007	0.329	0.219	0.463	0.335	0.650	0.560
IES	Income – STC	1995	0.286	0.269	0.434	0.415	0.622	0.605
		2000	0.429	0.302	0.559	0.440	0.710	0.621
		2005/2006	0.338	0.221	0.488	0.373	0.657	0.576
	Expenditure – STC	1995	0.300	0.278	0.447	0.423	0.629	0.610
		2000	0.430	0.292	0.564	0.441	0.714	0.624
		2005/2006	0.303	0.232	0.466	0.390	0.654	0.603
	Income - COICOP	1995	0.318	0.281	0.462	0.427	0.642	0.607
		2000	0.442	0.302	0.572	0.440	0.723	0.621
		2005/2006	0.316	0.221	0.473	0.379	0.652	0.586
	Consumption - COICOP	1995	0.339	0.189	0.502	0.341	0.691	0.566
		2000	0.458	0.198	0.601	0.343	0.753	0.560
		2005/2006	0.320	0.121	0.500	0.270	0.699	0.528
OHS	Expenditure – No imputations	1996	0.588	0.190	0.704	0.343	0.815	0.556
		1997	0.665	0.203	0.768	0.345	0.875	0.589
		1998	0.667	0.168	0.781	0.326	0.871	0.573
		1999	0.652	0.258	0.742	0.408	0.838	0.652
	Income – No imputations	1999	0.518	0.457	0.617	0.584	0.745	0.734
	Expenditure – After SRMI2	1996	0.565	0.204	0.687	0.337	0.816	0.545
		1997	0.660	0.205	0.764	0.374	0.870	0.608
		1998	0.656	0.207	0.771	0.345	0.865	0.578
		1999	0.634	0.273	0.727	0.442	0.829	0.648
	Income – After SRMI2	1999	0.494	0.512	0.596	0.622	0.729	0.756

[A] Without adjustments [B] With adjustments

Table 6.14: Continued

Survey	Per capita variable	Year	R211		R322		R593	
			[A]	[B]	[A]	[B]	[A]	[B]
LFS	Expenditure – No imputations	2001	0.693	0.324	0.773	0.476	0.859	0.683
		2002	0.684	0.373	0.788	0.515	0.853	0.713
		2003	0.678	0.456	0.758	0.555	0.838	0.745
		2004	0.649	0.415	0.738	0.599	0.827	0.738
	Expenditure – After SRMI2	2001	0.682	0.342	0.764	0.466	0.852	0.682
		2002	0.674	0.365	0.779	0.520	0.845	0.706
		2003	0.669	0.468	0.750	0.635	0.830	0.750
		2004	0.639	0.424	0.730	0.599	0.820	0.730
GHS	Expenditure – No imputations	2002	0.689	0.302	0.778	0.452	0.861	0.676
		2003	0.681	0.387	0.762	0.523	0.845	0.684
		2004	0.637	0.244	0.733	0.397	0.823	0.600
		2005	0.618	0.238	0.710	0.400	0.840	0.613
		2006	0.619	0.219	0.731	0.384	0.842	0.619
		2007	0.614	0.222	0.695	0.369	0.822	0.571
		2008	0.618	0.341	0.712	0.490	0.829	0.675
		2009	0.552	0.391	0.675	0.539	0.790	0.706
	Expenditure – After SRMI2	2002	0.677	0.321	0.768	0.449	0.854	0.674
		2003	0.668	0.377	0.751	0.510	0.837	0.707
		2004	0.627	0.243	0.723	0.410	0.815	0.627
		2005	0.612	0.234	0.705	0.440	0.836	0.612
		2006	0.615	0.217	0.728	0.381	0.839	0.615
		2007	0.611	0.220	0.692	0.366	0.820	0.567
		2008	0.610	0.373	0.706	0.528	0.824	0.676
		2009	0.549	0.386	0.674	0.536	0.790	0.705
PSLSD	Income	1993	0.475	0.346	0.598	0.474	0.745	0.648
	Expenditure	1993	0.398	0.180	0.566	0.346	0.750	0.589
NIDS	Income	2008	0.302	0.161	0.471	0.288	0.656	0.520
	Expenditure	2008	0.386	0.158	0.532	0.318	0.687	0.537
AMPS	Income	1993	0.438	0.278	0.586	0.438	0.737	0.643
		1994	0.439	0.264	0.593	0.420	0.735	0.632
		1995	0.464	0.291	0.594	0.434	0.741	0.621
		1996	0.473	0.277	0.610	0.437	0.744	0.636
		1997	0.456	0.264	0.589	0.407	0.732	0.607
		1998	0.453	0.269	0.583	0.415	0.725	0.603
		1999	0.469	0.259	0.591	0.415	0.723	0.599
		2000	0.458	0.273	0.582	0.428	0.723	0.609
		2001	0.466	0.280	0.579	0.425	0.717	0.598
		2002	0.434	0.246	0.563	0.387	0.709	0.573
		2003	0.418	0.241	0.554	0.388	0.704	0.583
		2004	0.415	0.206	0.548	0.362	0.703	0.560
		2005	0.391	0.205	0.519	0.345	0.680	0.530
		2006	0.385	0.193	0.512	0.328	0.673	0.516
		2007	0.332	0.161	0.455	0.298	0.613	0.470
		2008	0.292	0.169	0.410	0.283	0.580	0.451
		2009	0.306	0.162	0.414	0.282	0.574	0.437

The results above leads to the following question: Should the survey distributions should be adjusted, and should the post-adjustments per capita variables should be used to examine the real extent of poverty in South Africa? The analyses in this section have shown that the poverty trends were not affected significantly by the adjustments, but the levels of poverty were. With regard to the latter finding, this has serious implications on government's decisions on, for instance, the amount of each type of social grant to be given to the beneficiaries, expenditure on programs to create jobs, delivery of free basic services to households, etc. For instance, if the government thinks the number of poor under the poverty line as derived using the adjusted per capita variables was not as high as originally found (using the unadjusted per capita variables), perhaps this leads to the government to decide not to increase expenditure on social grants to alleviate poverty, but the tax revenue earned would rather be spent on other activities like infrastructure, education, etc.

6.9 Poverty and inequality trends revisited after applying the cross entropy approach

As discussed throughout the dissertation, the survey data under study are, strictly speaking, cross-sectional and not designed for time-series analyses because the weighting techniques are not the same across the surveys. Hence, in this section, the poverty and inequality trends in the two censuses and CS 2007, the three IESs, OHSs, LFSs and GHSs are re-visited after the application of the minimum cross entropy (CE) approach to re-weight these datasets[112], in order to find out if these trends would differ significantly after a consistent weighting technique is applied. Note that it was not impossible to re-weight the AMPSs, because as discussed in Section 2.8.1, the unique household number variable was not available, so it was impossible to identify members from the same household.

First, Tables 6.15 and Figures G.1-G.4 in Appendix G report the findings on the poverty headcount ratios, with the focus being the poverty headcount ratios using the R322 poverty line. Looking at the poverty trends in the two censuses and CS 2007, after re-weighting the latter survey by the CE approach, the poverty headcount ratios showed a negligible increase at all three poverty lines compared with the ratios using the original Stats SA weights (See Figure G.1). The same poverty trends were still observed, i.e., a moderate increase of poverty between the censuses, before a rapid decrease took place between Census 2001 and CS 2007.

[112] As the two censuses were not surveys, they were not re-weighted. In addition, since the QLFS took place during a 3-month period, the February population figure derived by the ASSA model was used to derive the CE weights in the Q1 survey. Similarly, the May, August and November ASSA model's population figures were used to derive the CE weights in the Q2, Q3 and Q4 surveys respectively. Since IES 2005/2006 was conducted between September 2005 and August 2006, the March 2006 population figure derived by the ASSA model was used to derive the CE weights for this survey. As NIDS took place between January and December 2008, the mid-year population figure derived by the ASSA model was used to derive the CE weights.

Table 6.15: Poverty headcount ratios at different poverty lines before and after the cross entropy approach was conducted, using the per capita variables

Survey	Per capita variable	Year	R211		R322		R593	
			[A]	[B]	[A]	[B]	[A]	[B]
Census/ CS	Income – No imputations	1996	0.501	0.501	0.606	0.606	0.728	0.728
		2001	0.568	0.568	0.670	0.670	0.789	0.789
		2007	0.397	0.403	0.529	0.534	0.694	0.696
	Income – After SRMI1	1996	0.493	0.493	0.601	0.601	0.726	0.726
		2001	0.547	0.547	0.647	0.647	0.769	0.769
		2007	0.351	0.357	0.478	0.484	0.656	0.659
	Income – After SRMI2	1996	0.441	0.441	0.576	0.576	0.715	0.715
		2001	0.447	0.447	0.592	0.592	0.750	0.750
		2007	0.329	0.336	0.463	0.469	0.650	0.652
IES	Income – STC	1995	0.286	0.297	0.434	0.445	0.622	0.630
		2000	0.429	0.430	0.559	0.557	0.710	0.707
		2005/2006	0.338	0.333	0.488	0.479	0.657	0.647
	Expenditure – STC	1995	0.300	0.310	0.447	0.457	0.629	0.637
		2000	0.430	0.429	0.564	0.561	0.714	0.711
		2005/2006	0.303	0.298	0.466	0.457	0.654	0.644
IES	Income - COICOP	1995	0.318	0.330	0.462	0.472	0.642	0.650
		2000	0.442	0.442	0.572	0.570	0.723	0.719
		2005/2006	0.316	0.312	0.473	0.464	0.652	0.643
	Consumption - COICOP	1995	0.339	0.350	0.502	0.514	0.691	0.701
		2000	0.458	0.458	0.601	0.599	0.753	0.750
		2005/2006	0.320	0.315	0.500	0.493	0.699	0.690
OHS/ LFS	Expenditure – No imputations	1996	0.588	0.602	0.704	0.722	0.815	0.836
		1997	0.665	0.650	0.768	0.755	0.875	0.867
		1998	0.667	0.657	0.781	0.774	0.871	0.867
		1999	0.652	0.647	0.742	0.736	0.838	0.834
		2001	0.693	0.683	0.773	0.765	0.859	0.854
		2002	0.684	0.675	0.788	0.780	0.853	0.849
		2003	0.678	0.669	0.758	0.750	0.838	0.833
		2004	0.649	0.642	0.738	0.731	0.827	0.821
	Expenditure – After SRMI2	1996	0.565	0.578	0.687	0.701	0.816	0.830
		1997	0.660	0.645	0.764	0.751	0.870	0.862
		1998	0.656	0.647	0.771	0.765	0.865	0.861
		1999	0.634	0.629	0.727	0.721	0.829	0.825
		2001	0.682	0.672	0.764	0.756	0.852	0.847
		2002	0.674	0.664	0.779	0.770	0.845	0.840
		2003	0.669	0.659	0.750	0.742	0.830	0.824
		2004	0.639	0.631	0.730	0.722	0.820	0.814
GHS	Expenditure – No imputations	2002	0.689	0.686	0.778	0.772	0.861	0.856
		2003	0.681	0.681	0.762	0.758	0.845	0.841
		2004	0.637	0.626	0.733	0.722	0.823	0.816
		2005	0.618	0.609	0.710	0.701	0.840	0.834
		2006	0.619	0.611	0.731	0.723	0.842	0.836
		2007	0.614	0.607	0.695	0.687	0.822	0.817
		2008	0.618	0.615	0.712	0.708	0.829	0.826
		2009	0.552	0.564	0.675	0.683	0.790	0.795
	Expenditure – After SRMI2	2002	0.677	0.674	0.768	0.762	0.854	0.848
		2003	0.668	0.668	0.751	0.747	0.837	0.832
		2004	0.627	0.616	0.723	0.712	0.815	0.807
		2005	0.612	0.603	0.705	0.696	0.836	0.831
		2006	0.615	0.607	0.728	0.719	0.839	0.833
		2007	0.611	0.603	0.692	0.684	0.820	0.815
		2008	0.610	0.607	0.706	0.702	0.824	0.821
		2009	0.549	0.560	0.674	0.682	0.790	0.795

[A]: Stats SA weights [B]: Cross entropy weights

With regard to the poverty trends in the IESs, after using the CE weights, the poverty headcount ratio increased slightly in IES 1995, but the opposite took place in IES 2000 and IES 2005/2006 (See Figure G.2). However, the same poverty trends were still observed, i.e., a rapid increase between 1995 and 2000, before it decreased between the 2000 and 2005/2006 IESs, but the IES 2005/2006 poverty headcount ratios were higher than the IES 1995 ratios. Next, looking at OHS 1996-1999 and the 2001-2004 September LFSs, Figure G.3 shows that the use of the CE weights resulted in slightly lower poverty headcount ratios in all surveys, except in OHS 1996. On the other hand, the poverty headcount ratios in GHS 2002-2009 also experienced a slight decrease in all surveys after using the CE weights, except in GHS 2009 (See Figure G.4). In addition, the use of the CE weights did not cause any changes in the poverty trends in the OHSs, LFSs and GHSs in general.

Finally, Table 6.16 presents the Gini coefficients using the Stat SA weights and CE weights respectively, and the results show that the Gini coefficients only showed very negligible changes after using CE weights, and the inequality trends did not show any changes.

Table 6.16: Gini coefficients, before and after the cross entropy approach was conducted

Survey	**Per capita variable**	**Year**	**Gini coefficient**	
			[A]	**[B]**
Census/CS	Income – No imputations	1996	0.742	0.742
		2001	0.825	0.825
		2007	0.774	0.774
	Income – After SRMI1	1996	0.734	0.734
		2001	0.817	0.817
		2007	0.759	0.761
	Income – After SRMI2	1996	0.694	0.694
		2001	0.756	0.756
		2007	0.743	0.745
IES	Income – STC	1995	0.655	0.653
		2000	0.711	0.716
		2005/2006	0.717	0.716
	Expenditure – STC	1995	0.660	0.658
		2000	0.710	0.713
		2005/2006	0.733	0.729
	Income - COICOP	1995	0.660	0.659
		2000	0.709	0.714
		2005/2006	0.716	0.715
	Consumption - COICOP	1995	0.612	0.610
		2000	0.651	0.656
		2005/2006	0.670	0.670
OHS/LFS	Expenditure – No imputations	1996	0.646	0.651
		1997	0.663	0.670
		1998	0.662	0.662
		1999	0.713	0.709
		2001	0.745	0.740
		2002	0.781	0.778
		2003	0.813	0.816
		2004	0.815	0.815

Table 6.16: Continued

Survey	Per capita variable	Year	Gini coefficient	
			[A]	[B]
OHS/LFS	Expenditure – After SRMI2	1996	0.636	0.650
		1997	0.660	0.667
		1998	0.659	0.659
		1999	0.702	0.701
		2001	0.739	0.736
		2002	0.779	0.778
		2003	0.821	0.819
		2004	0.815	0.815
GHS	Expenditure – No imputations	2002	0.736	0.740
		2003	0.772	0.777
		2004	0.720	0.718
		2005	0.737	0.736
		2006	0.753	0.752
		2007	0.735	0.738
		2008	0.787	0.788
		2009	0.815	0.816
	Expenditure – After SRMI2	2002	0.736	0.740
		2003	0.771	0.776
		2004	0.723	0.721
		2005	0.738	0.737
		2006	0.748	0.747
		2007	0.735	0.737
		2008	0.787	0.788
		2009	0.806	0.807

[A]: Stats SA weights [B]: Cross entropy weights

To conclude, after re-weighting the datasets to re-examine the poverty headcount ratios, the use of the CE weights did not result in any major changes in the poverty trends since the transition. This implies that the abrupt changes observed in some surveys cannot be attributed to inconsistent Stats SA weights, but rather are real or are influenced by other factors like the questionnaire design, as well as the way the income and expenditure information was captured, as discussed in Chapter 3 (e.g., recall method vs. diary method, single-estimate approach vs. aggregation approach, etc.).

6.10 Chapter summary

This chapter began by giving a literature review of recent studies on the South African poverty and inequality trends. It was found that, in general, poverty increased since the transition, before a continuous downward trend took place since 2000. In addition, inequality worsened since the transition until 2000, before rising further in the 2000s, but the extent of such increase was not as rapid as what happened in the 1990s.

Next, poverty trends using the surveys under study were investigated by focusing on the FGT poverty indices. It was found that, although the poverty levels differed amongst the surveys, similar poverty trends were generated, i.e., increase of poverty headcount ratios from 1994 until 1999, before a downward trend was observed since 2000. These results were consistent with what was found in the recent studies. In addition, the poverty headcount ratios were relatively higher in OHSs/LFSs and GHSs. Also, the ratios decreased (especially in the two censuses and CS 2007) after SRMI was conducted. The factors discussed in Chapter 3 (e.g., number of intervals and width of each interval, the use of the diary method to complement recall method, and so forth) might have played a role to cause the poverty levels to be different amongst the surveys.

Next, inequality trends were analysed by looking at the Gini coefficients as well as Theil-L and Theil-T indices. With regard to Gini coefficients, the surveys gave different trends, e.g., the AMPSs showed a very stable trend, the IESs, OHSs/LFSs and GHSs displayed an upward trend, while the censuses and CS 2007 showed that the Gini coefficient first increased before decreasing. In addition, the inequality decomposition showed that within-race inequality as a share of total inequality increased and became more dominant since the transition. Hence, the inequality trends across the surveys were not as comparable as the poverty trends.

As NIDS is the only survey that captured information on total income and expenditure amounts by means of both the single-estimate and aggregation approaches, the poverty headcount ratios and Gini coefficients derived from using these two approaches were looked at, and in Section 6.5, it was found that the values of these measures became greater when using the single-estimate approach. The results implied that the income and expenditure estimates and the subsequent poverty and inequality aggregates might be more precise and reliable if the aggregation approach is adopted in the survey.

The last three sections of Chapter 6 aimed at improving the reliability and comparability of poverty and inequality estimates across the surveys further. In Section 6.7, it was found that when the AMPS 2000 intervals and midpoint-Pareto method were applied on the IES 2000 data, it resulted in very similar poverty and inequality estimates as those derived by using the continuous data as it was. Hence, the midpoint-Pareto was applied in all three IESs to re-examine the poverty and inequality trends. In addition, since there is a lack of South African studies on the effect the number and width of intervals have on poverty and inequality estimates, different intervals (i.e., in addition to AMPS 2000 intervals, census intervals, GHS intervals, etc. were also applied) were used in the three IESs. The results showed that the use of fewer and wider intervals did not necessarily result in greater inequality as found by Seiver (1979). In contrast, the same poverty trends (i.e., an increase between IES 1995 and IES 2000, before a decline took place between IES 2000 and IES 2005/2006) were observed when applying these intervals, but the use of the relatively wider Census 2001 intervals resulted in lower poverty headcount ratios in all three IESs. The AMPS 2007, CS 2007 and GHS 2007 intervals were also applied on the single-estimated and aggregated income and expenditure variables in NIDS, and it was found that poverty headcount ratios became lower but Gini coefficients showed quite a big increase, after application of the GHS bands. This result implies that the use of fewer intervals lead to greater inequality, confirming Seiver's findings.

In Sections 6.8 and 6.9, poverty and inequality trends were investigated again by shifting the survey distribution rightwards in line with the national accounts mean and using the CE weights (instead of the original Stats SA weights) respectively. As expected, it was found that the former approach resulted in lower poverty headcount ratios during the period under investigation, but there was no change in the poverty trends. On the other hand, the poverty headcount ratios and Gini coefficients showed negligent changes after using the CE weights, implying that the abrupt changes in these estimates in some surveys were not attributed to possible errors in the post-stratified Stats SA weights.

To conclude, the results of the analyses in Chapter 6 found that it is not easy to determine which survey provided the most reliable poverty and inequality levels and trends, as these estimates could be influenced by the various factors as discussed in Chapter 3, thereby making the comparisons across the surveys difficult. Nonetheless, it is certain that the adoption of midpoint-Pareto method to derive the continuous amount in each interval (if the income and/or expenditure information was captured in bands) is reliable; that it is necessary to deal with households with zero or unspecified income by means of an appropriate imputation method to improve the reliability of poverty and inequality estimates (especially for surveys with high proportion of households reporting zero or unspecified income); that the comparability of poverty and inequality trends across the surveys was not significantly improved by the use of a consistent weighting technique; and there is no indication that expenditure was captured better by the surveys and the subsequent poverty and inequality estimates were more reliable.

CHAPTER SEVEN: CONCLUSION

7.1 Introduction

Numerous sources of survey data were used in this dissertation to investigate money-metric poverty and inequality as well as labour market trends in South Africa since the transition. OHSs/LFSs/QLFSs were the most commonly used data to examine the labour market trends, as these were the surveys with the primary aim of capturing labour market information. With regard to the poverty and inequality trends, the IES, census and CS 2007 data were used in most of the studies. Moreover, the Stellenbosch University economics researchers recently used the AMPS data for these analyses.

However, the surveys are different in various aspects, such as sample size, sampling design, questionnaire design, labour market status derivation methodology, the way the questions relating to income and expenditure were asked, weighting techniques, etc. Hence, this dissertation investigated these issues, before analysing the post-apartheid labour market, poverty and inequality levels and trends using all the available survey data, as well as the possible factors accounting for the differences and these levels and trends, if any, amongst the surveys. These factors were addressed, if possible, in order to improve the comparability of the data across time and between different surveys and censuses, before the labour market, poverty and inequality trends were examined again.

7.2 Review of findings

This dissertation began by reviewing the South African survey data between 1993 and 2009 in Chapter 2, namely PSLSD 1993, OHS 1994-1999, LFS 2000-2007, QLFS 2008-2009, Census 1996 & 2001, CS 2007, GHS 2002-2009, NIDS 2008 as well as AMPS 1993-2009. The focus of the analysis was on the sampling design, sample size, questionnaire structure, derivation of labour market status, as well as how the information on income and expenditure (consumption) was captured.

OHS, LFS and QLFS were the surveys with the primary aim of capturing labour market information (and hence were used for examining labour market trends in Chapter 4), but it was found that, due to the absence of the metadata document in OHS 1994-1995, the labour market status derivation methodology in this survey was not known. In addition, there were slight changes in the methodologies to derive both the narrow and broad status throughout the years. Furthermore, the broad labour market status derivation methodologies before and after the introduction of the QLFS were incomparable.

With regard to the income and expenditure information, the analyses in Chapter 2 found that it was captured differently when comparing the various surveys. Some surveys only captured income, others only expenditure, and only the three IESs, OHS 1999, PSLSD and NIDS collected both income and expenditure information. In addition, respondents were asked to report the actual amounts in some surveys, but declare the relevant intervals in other surveys.

If the respondents reported the actual amounts, they were either asked to declare a straightforward, 'one-shot' single estimate amount (e.g., OHS 1996-1998, NIDS) or were first asked in detail the amount spent on each income or expenditure item, before the aggregate income or expenditure was derived (e.g., IESs, PSLSD, NIDS). In contrast, if the survey participants were asked to report the relevant income or expenditure category, the number of intervals as well as the width of intervals differed a lot amongst the surveys. For instance, there were only eight intervals in the GHSs, but as many as 32 in AMPS; the width of the interval was as narrow as R100 in AMPS but as wide as R102 400 in Census 2001 and CS 2007. Finally, it was found that the proportion of households reporting zero or unspecified income or expenditure was relatively high in the two censuses and CS 2007.

In Chapter 3, various issues that could affect the reliability of the labour market, poverty and inequality trends were investigated. First, although the general conclusion from the literature was that expenditure is the preferred variable to be used for poverty and inequality analyses for a developing country like South Africa, a more in-depth review of the literature showed that this is not always the case. In fact, using income or expenditure variable involves its pros and cons. Secondly, in all surveys under study, the recall method was used to collect income and expenditure information, except in IES 2005/2006 which used both the recall and diary methods to capture the information. Each method has its advantages and disadvantages. It was also found that durable expenditure would always be under-estimated to some extent, regardless of which method was used.

In some surveys the respondents were asked to declare income and/or expenditure in actual amounts, while in other surveys they were asked to report the relevant interval they fell in. Looking at the former approach in greater detail, the participants were asked to declare the 'one-shot', single estimate total income or expenditure amount, or were asked a series of questions on how much they earned or spent in each source, before the total amount was derived (i.e., the aggregation approach). Both merits and drawbacks are involved in these two approaches.

If the participants were asked to declare the relevant interval, three issues came up. First, the appropriate method to use to approximate the amount in each band, i.e., making the data continuous before the per capita income or expenditure variable could be derived for the subsequent poverty and inequality analyses. Various methods (midpoint method, mid-point Pareto method, interval regression, random midpoint method, and equal distribution method) were discussed, and the review of the literature found that it was acceptable to apply the midpoint methods in all intervals, except in the open interval, where the Pareto method had to be applied. This was referred to as the midpoint-Pareto method throughout the dissertation.

With regard to the second issue, namely the impact of the number and width of the bands on poverty and inequality estimates, it was found that there was lack of both South African and international studies done on this issue. Thirdly, various methods to deal with households with zero or unspecified income or expenditure were dealt with, such as casewise deletion, available-case deletion, single imputation and multiple imputation, and it was decided to use a multiple imputation approach, namely sequential regression multiple imputation (SRMI) for poverty and inequality analyses in Chapters 5 and 6.

Furthermore, as the surveys in general under-estimated income or expenditure when comparing with the national accounts total income of the same year, it has been argued that the survey distribution should be adjusted rightwards in line with the national accounts income mean, before the poverty and inequality trends are examined. However, the other side of the argument is that very different methodologies were involved to capture income in surveys and national accounts, hence it might not be correct to simply shift the survey distribution; perhaps the problem comes from the possibly unreliable survey distribution. Hence, Chapter 3 reviewed the existing literature on the arguments for and against such adjustments.

Chapter 3 also considered the entropy approach to re-weight the survey data, as the data was, strictly speaking, cross-sectional and not time-series data, due to the inconsistent post-stratification weighting techniques adopted by Stats SA. Hence, the entropy post-stratification approach was adopted to re-weight the person weights of all the data under study by conforming to the racial, gender, age and provincial distribution of the population estimates as derived by the ASSA 2003 model. The method was discussed in detail in the chapter, and it was concluded that this approach to re-weight the data is more time consistent and would be used to re-investigate labour market, poverty and inequality trends in Chapters 4-6.

Chapter 3 then investigated the changes in the labour market status derivation methodology in the OHSs, LFSs and QLFSs in greater detail, and these changes could lead to some abrupt changes in the labour market aggregates since the transition. Finally, other factors that could also influence the comparability and reliability of the labour market, poverty and inequality aggregates across the surveys were examined, such as the length of the questionnaire, whether the questions to capture the information to derive these aggregates were asked early or towards the end of the questionnaires, and possible fatigue happening to both interviewers and interviewees.

Chapter 4 began by reviewing the recent studies on the South African labour market trends. It was found that there have been three types of studies on these trends since the transition: (1) studies that compared OHS 1995 with the latest available OHS or LFS at the time of writing to derive the 'trends' on labour force size (LF), labour force participation rates (LFPRs), employment, unemployment and unemployment rates; (2) studies that used most of or all available datasets at the time of writing to examine the same trends over a period of time; (3) studies that also used most of or all available datasets, but the focus was rather on wage trends, as well as the employment gap and wage gap between groups (e.g. by gender or race). The literature review found that the unemployment rate remained high since the political transition, despite increased employment, because the pace of increase of job creation was not rapid enough to absorb the expanding labour force. In addition, blacks and females were more likely to be unemployed.

Chapter 4 next investigated what happened to the LF, LFPRs and employment, work activities of the employed, as well as unemployment and unemployment from 1994 to 2009 by using all available OHSs, LFSs and QLFSs during the period. As it was earlier found that the broad labour market status derivation methodologies before and after the introduction of QLFS were incomparable, the focus of the analysis of Chapter 4 was on the narrow terms. It was found that, in general, the LF and LFPR experienced a rapid increase during the OHSs, followed by an abrupt increase during the changeover from OHS to LFS. The narrow LF and

LFPR have since increased slightly, while the broad LF and LFPR have stabilised. The trends in the LFSs and QLFSs did not suggest that any feminisation of the labour force had taken place after the OHS years. In fact, males (and blacks) were more likely to have become labour force participants.

In addition, the number of employed showed huge fluctuations, and it was only since LFS 2004b that employment increased in a stable and continuous fashion. Hence, if different reference points are used in the calculation of target growth ratio (TGR), actual growth ratio (AGR) and employment absorption ratio (EAR), one may draw contradictory conclusions regarding whether job creation or jobless growth occurred place in the South African economy. It was also found that both the narrow and broad unemployment rates increased continuously from OHS 1994 to LFS 2003a, followed by a continuous downward trend from LFS 2003b onwards until an upward trend was observed from 2009 due to the impact of recession. Furthermore, blacks, females, those in the younger age cohorts (i.e., 15-24 years and 25-34 years at the time of survey), as well as those without Matric were more likely to be unemployed.

The last section of Chapter 4 aimed at addressing the issues discussed in Chapter 3 to improve the comparability of labour market aggregates across the surveys. First, it was found that the non-agricultural formal employment was very comparable with the SEE and QES data, and the extent of the abrupt changes as observed previously (when informal sector employment, agricultural employment and self-employed were included in the analyses, as in Sections 4.4) became much less serious. Next, the labour market trends were re-examined by using the minimum cross entropy weights, and it was found that these trends (especially the abrupt changes observed in certain surveys when using the original Stats SA weights) remained the same. This implies that the discrepancies in the labour market trends were not caused by the inconsistent Stats SA weights, but rather by the better capturing of labour market information due to the improvement of the questionnaire design, the actual increase of labour force and employment since the transition, or due to the drastic changes in the labour market status derivation methodology across the surveys.

Chapter 4 concluded by applying the LFS 2000b-LFS 20007b labour market status derivation methodology in OHS 1996-LFS 2000a, and it was found that the abrupt changes in the broad unemployment between OHS 1999 and LFS 2000a no longer happened. In contrast, when the revised QLFS methodology was applied in all LFSs and QLFSs, the number of broad unemployed became lower in the LFSs (despite still being relatively greater when compared with the QLFSs), and the extent of the sudden decrease of broad unemployed became less serious. Therefore, the results suggest that the comparability of the labour market aggregates across the surveys could be improved to a certain extent by applying a consistent labour market status derivation methodology in all surveys.

Chapter 5 began by reviewing poverty and inequality concepts and measurements. In addition, the SRMI approach to impute zero or unspecified income or expenditure in the surveys under study was discussed in greater detail. The derivation of real per capita income, expenditure and consumption variables was explained, and the comparison of the survey income / expenditure / consumption amounts with the national accounts income data was examined. It was found that the post-SRMI income variable in OHS 1999 was the only survey that reported a higher amount than the national accounts income from the same year. That is, other surveys captured lower income / expenditure / consumption amounts when compared with national accounts income data. The under-estimation was particularly more serious in the OHSs, LFSs and GHSs. It was also found that there was a rapid drop of income and expenditure between IES 1995 and IES 2000.

A literature review of recent South African studies on money-metric poverty and inequality trends was the focus of the beginning of Chapter 6. Per capita variables were used in some studies while others used per adult equivalent variables. Also, households with zero or unspecified income or expenditure were simply excluded from the analyses in some studies but had the amounts imputed in others. In addition, poverty lines used in the studies differed, but the three proposed poverty lines of Woolard and Leibbrandt (2006) were commonly used in recent studies. Despite the aforementioned differences, the general conclusion from the literature review was that the poverty headcount ratio increased moderately between 1994 and 2000, before a strong downward trend took place since 2000. With regard to inequality, the Gini coefficient increased moderately in some studies but the increase was stronger in others between 1994 and 2000, though the coefficient stabilised since 2000. Furthermore, within-race inequality as a share of total inequality increased continuously and became the dominant component since the transition.

Chapter 6 next used the pre-SRMI and post-SRMI per capita variables derived in Chapter 5 as well as the three poverty lines proposed by Woolard and Leibbrandt to examine poverty trends since the transition, focusing exclusively on the Foster-Greer-Thorbecke (FGT) measures. It was found that poverty increased since the transition, before declining since 2000. The only exception was AMPS, which found that the poverty level was quite stable between 1993 and 1999. Despite the fact that, in general, the poverty trends found were consistent with the results of the recent studies, the levels of poverty differed a lot in various surveys. For instance, the poverty levels were clearly higher in the OHSs, LFSs and GHSs (which involved fewer intervals), and as expected, poverty levels decreased after the application of SRMI in the two censuses and CS 2007. These differences across the surveys could be attributed to the factors discussed in Chapter 3.

With regard to inequality trends as measured by the Gini coefficients, the surveys gave contrasting results in both the levels and trends. Using OHSs, LFSs, GHSs and IESs, it was found that the Gini coefficient increased in general throughout the years, but it was extremely stable in AMPSs. In contrast, measured inequality increased between the two censuses but decreased between Census 2001 and CS 2007. In addition, the Gini coefficients were relatively smaller in AMPSs (which involved more and narrow intervals) and IESs (which collected income and expenditure information by means of the aggregation approach), but greater in the two censuses and CS 2007 (these surveys were characterised by very wide income intervals) and the GHSs. Furthermore, in all surveys under study, it was found that within-race inequality as a share of total inequality became more prominent. Hence, it was once again likely that the factors examined in Chapter 3 might have played a role to cause the poverty and inequality levels and trends to differ across the surveys, thereby affecting the comparability and reliability of these estimates.

The last few sections of Chapter 6 tried to address some of these factors in greater detail to see if the comparability and reliability of these estimates would improve. First, the NIDS poverty and inequality estimates in greater detail. This survey was the only one that captured income and expenditure information using both the single estimate and aggregation approaches, and it was found that the single estimate income and expenditure amounts were significantly lower when compared with the aggregate amounts. Hence, as expected, the poverty levels were higher when using the single estimate variables. It was also found that the Gini coefficients were higher when using these variables.

Chapter 6 moved on to apply the AMPS 2000 intervals on the IES 2000 income data, before using various methods to make the categorical data continuous again. The FGT indices and Gini coefficients were derived, and it was found that these estimates using the midpoint-

Pareto method as well as those derived from the actual continuous income variable were very similar. Hence, this confirmed the findings of recent studies that it is safe to use the midpoint-Pareto approach. Next, the chapter investigated the impact of the number and width of intervals on the poverty and inequality estimates in IES 2000. First, the AMPS 2000 intervals were collapsed in various ways before they were applied on the IES data, and it was found that the poverty and inequality estimates only showed negligent changes. Next, the Census 1996, Census 2001, GHS 2009, as well as R500, R1 000 and R2 000 intervals in 2000 prices were also applied on the IES 2000 income data, and the results of the analyses showed that the poverty indices and Gini coefficient were lower when using the R2 000 intervals.

These intervals were also applied in the IES 1995 and IES 2005/2006 income data, and it was found that the poverty trends remained the same (i.e., poverty increased between 1995 and 2000, before a decreased took place). In addition, after the application of the AMPS 2000, Census 1996 and Census 2001 intervals, the same trend as in using the IES actual continuous income variable was found, i.e., the Gini coefficient increased rapidly between 1995 and 2000, before a slight increase took place between IES 2000 and IES 2005/2006. However, after using the GHS 2009, R500, R1 000 and R2 000 intervals, although the Gini coefficient once again increased between the first two IESs, it decreased between the last two IESs.

The next experiment on investigating the impact of the number and width of intervals on poverty and inequality estimates was to apply the AMPS 2007, CS 2007 and GHS 2007 intervals (in 2000 prices) on the NIDS single estimate income and expenditure data. It was found that the poverty levels were very similar (when compared with the actual continuous income and expenditure variables), but the Gini coefficient was much higher when using the GHS 2007 intervals. This result conformed with the finding of Seiver (1979) that the use of fewer intervals leads to increased inequality estimates.

The abovementioned results must still be interpreted with caution, as it was assumed that the respondents who declared their income or expenditure by the aggregation approach in the IESs and NIDS, would report similar income or expenditure if asked to report ‘one-shot’, categorical answers. However, this is not always the case, as already discussed in Sections 3.5 and 6.6.

Finally, Chapter 6 investigated poverty and inequality trends by shifting the survey distribution rightwards in line with the national accounts mean as well as using the CE weights (instead of the original Stats SA weights). As expected, it was found that the former approach resulted in lower poverty headcount ratios during the period under investigation, but there was no change in the poverty trends. In contrast, the poverty headcount ratios and Gini coefficients showed negligent changes after using the CE weights, implying that the abrupt changes in these estimates in some surveys were not attributed to possible errors in the post-stratified Stats SA weights.

7.3 Conclusion

The empirical analyses in the dissertation found that, in general, poverty increased since the transition, before it decreased since 2000. However, different inequality trends were found when using various surveys. Also, the labour market aggregates showed very abrupt fluctuations in the OHSs, during the changeover between OHS and LFS in both narrow and broad terms, and again between LFS 2007b and QLFS 2008Q1 in broad terms. The results highlight various issues that deserve attention:

- One cannot blindly use the survey data for conducting labour market, poverty and inequality analyses, but must interpret these trends with caution, as the surveys are very different in many aspects (as discussed in Chapters 2 and 3).
- It is not possible to derive comprehensive labour market trends by only comparing OHS 1995 (the labour market status derivation methodology of this survey was not known, due to the absence of the metadata document) and an OHS, LFS or QLFS. All surveys should be considered when deriving the trends.
- Even if all labour surveys were taken into consideration, it does not mean that the labour market trends derived are fully reliable and comparable. In fact, the abrupt changes in the labour market aggregates across certain surveys could be identified more clearly by examining all surveys.
- The QLFS labour market status derivation methodology and the OHS/LFS methodologies are really not comparable to a certain extent, thereby causing very different unemployment aggregates in broad terms. Applying the QLFS methodology in the earlier surveys did improve the comparability of the results across the surveys to a certain extent. Unfortunately, the OHS/LFS and QLFS broad unemployed and unemployment rates are still not fully comparable.
- The abrupt labour market trends in the OHSs as well as during the changeover from OHS to LFS did not seem to be the result of inconsistent Stats SA weights, but rather due to the fact that these changes were real or could be attributed to the improved

questionnaire design to capture the labour market information better. The poverty and inequality trends were also not strongly influenced by the fact that the weighting techniques differed amongst the surveys.

- Whether the diary method is a better approach to collect the income and expenditure information would only be known if this approach is used more regularly in surveys.
- The survey data should be validated against external sources like national accounts income data and other sources like National Treasury's social grants expenditure and income tax revenue. Furthermore, as surveys tend to under-capture both income and expenditure, there might be a need to adjust the survey income (expenditure) distribution rightwards to be consistent with the national accounts series, in order to improve the validity and reliability of the poverty levels and trends across surveys.
- Poverty and inequality estimates were influenced by households reporting zero or unspecified income or expenditure. Simply accepting and including the zero-income households, or excluding zero-income and missing-income households from the analyses would affect the validity of these estimates as well as the comparability of these estimates across the surveys over the years.
- More research needs to be done to investigate the impact of the number and width of bands on the poverty and inequality estimates and trends.
- It is difficult to know which surveys provide the most reliable poverty and inequality estimates, as the surveys were already incomparable to a certain extent (Chapter 3). Even after addressing some of the factors affecting the poverty and inequality estimates (Chapter 6), the comparability of these estimates across surveys only improved to a certain extent. Thus, this raises the question of whether the income and expenditure information should be captured similarly across the surveys. For instance, should the number of intervals and the width of each interval be the same in next census and GHS, should there be more intervals with narrow range in each band in these censuses and surveys as in AMPSs, and should the level of disaggregation of income and expenditure items be the same in IESs and NIDS?
- As the income and expenditure information were collected so differently in each survey, the levels of poverty and inequality could differ a lot across the surveys. Yet, there is still a need to undertake the sort of analyses as done in this dissertation in order to make valid comparisons of both the poverty and inequality levels and trends across the surveys.

REFERENCES

Actuarial Society of South Africa. *ASSA 2003 full model*.
[Online] Available: http://aids.actuarialsociety.org.za/default.asp?pageid=3282

Actuarial Society of South Africa. *Note to the users of ASSA 2003 model*.
[Online] Available: http://aids.actuarialsociety.org.za/scripts/buildfile.asp?filename=Note to the users of Assa 2003 model.doc

African National Congress (1994). *A basic guide to the Reconstruction and Development Programme*. [Online] Available: http://www.anc.org.za/234

Agüero, J., Carter, M.R. and May, J. (2005). *Poverty and inequality in the first decade of democracy: Evidence from KwaZulu-Natal*. Paper presented at the "South African economic policy under democracy: A ten-year review" seminar, Stellenbosch.

Ahmed, N., Brzozowski, M. and Crossley, T.F. (2005). *Measurement errors in recall food expenditure data*. SEDAP Research Paper No. 133. Hamilton: The Program for Research on Social and Economic Dimensions of an Aging Population (SEDAP).

Ahmed, N., Brzozowski, M. and Crossley, T.F. (2006). *Measurement errors in recall food consumption data*. IFS Working Paper W06/21. London: Institute for Fiscal Studies.

Altman, M. (2003). *Jobless or job-creating growth? Some preliminary thoughts*. Paper presented at the DPRU/TIPS Forum, Johannesburg.

Altman, M. (2008). Revisiting South African employment trends in the 1990s. *South African Journal of Economics*. 76(2): 126 – 147.

Ardington, C., Lam, D., Leibbrandt, M. and Welch, M. (2005). *The sensitivity of estimates of post-apartheid changes in South African poverty and inequality to key data imputations*. CSSR working paper no. 106. Cape Town: Centre for Social Science Research.

Ardington, C., Lam, D., Leibbrandt, M. and Welch, M. (2006). The sensitivity to key data imputations of recent estimates of income poverty and inequality in South Africa. *Economic Modelling*. 23: 822-835.

Argent, J., Finn, A., Leibbrandt, M. and Woolard, I. (2009). *Poverty: Analysis of the NIDS wave 1 dataset*. Discussion Paper No. 13. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

Armstrong, P. and Burger, C. (2008). *Poverty, inequality and the role of social grants: An analysis through decomposition techniques*. Paper presented at the DPRU Conference, Muldersdrift.

Armstrong, P., Lekezwa, B. and Siebrits, K. (2008). *Poverty in South Africa: A profile based on recent household surveys*. Stellenbosch Economic Working Papers: 04/2008. Stellenbosch: Stellenbosch University.

Armstrong, P. and Steenkamp, J. (2008). *South African trade unions: An overview for 1995 to 2005*. Stellenbosch Economic Working Papers: 10/2008. Stellenbosch: Stellenbosch University.

Arora, A. and Ricci, L. A. (2005). Unemployment and the labour market. In Nowak M. and Ricci, L.A. (ed.): *Post-apartheid South Africa: The first ten years*. Washington: International Monetary Fund: 23-47.

Atkinson, A.B. (1970). On the measurement of inequality. *The Journal of Economic Theory*. 2(3): 244-263.

Banejee, A., Galliani, S., Levinsohn, J. and Woolard, I. (2006). *Why has unemployment risen in the new South Africa?* CID Working Paper No. 134. Harvard: Centre for International Development, Harvard University.

Barker, F. (2008). *The South African labour market*. 5th edition. Pretoria: Van Schaik Publishers.

Battinstin, E. (2003). *Errors in survey reports of consumption expenditures*. IFS Working Papers W03/07. London: Institute for Fiscal Studies.

Bhalla, S. (2002). *Imagine there's no country: Poverty, inequality, and growth in the era of globalization*. 1st edition. Washington: Peterson Institute.

Bhalla, S. (2010). Raising the standard: The war on global poverty. In Anand, S., Segal, P. and Stiglitz, J.E. (ed.), *Debates on the measurement of global poverty*. 1st edition. New York: Oxford University Press: 115-142.

Bhorat, H. (2003). *Employment, earnings and vulnerability in the South African labour market: An empirical investigation based on official survey data*. Unpublished PhD thesis. Stellenbosch: Stellenbosch University.

Bhorat, H. (2004). Labour market challenges in the post-apartheid South Africa. *South African Journal of Economics*. 72(5): 940 – 977.

Bhorat, H. (2005). *Changing patterns of employment and employer-employee relations in post-apartheid South Africa*. Paper presented at the “South African economic policy under democracy: A ten-year review” seminar, Stellenbosch.

Bhorat, H. (2006). Labour supply and demand constraints on employment creation: A microeconomic analysis. In Padayachee, V. (ed.), *The development decade? Economic and social change in South Africa: 1994 – 2004*. 1st edition. Pretoria: Human Sciences Research Council: 275 – 301.

Bhorat, H. (2009). *Unemployment in South Africa: Descriptors and determinants*. A paper presented at the Fourth IZA/World Bank Conference on Employment and Development, Bonn, Germany.

Bhorat, H. and Hinks, T.J. (2005). *Changing patterns of employment and employer-employee relations in post-apartheid*. Paper presented at the “South African economic policy under democracy: A ten-year review” seminar, Stellenbosch.

Bhorat, H. and Leibbrandt., M. (2001). Modelling vulnerability and low earnings in the South African labour market. In Bhorat, H., Leibbrandt, M., Maziya, M., Van der Berg, S. and Woolard, I. (ed.), *Fighting poverty: Labour markets and inequality in South Africa*: 107 – 129.

Bhorat, H., Naidoo, P. and Van der Westhuizen, C. (2006). *Shifts in non-income welfare in South Africa: 1993 – 2004*. DPRU Working Paper 06/108. Cape Town: Development Policy Research Unit, University of Cape Town.

Bhorat, H. and Oosthuizen, M. (2005). *The post-apartheid South African labour market*. DPRU Working Paper 05/93. Cape Town: Development Policy Research Unit, University of Cape Town.

Bhorat, H., Poswell, L. and Naidoo, P. (2004). *Dimensions of poverty in post-apartheid South Africa*. Cape Town: Development Policy Research Unit, University of Cape Town.

Bhorat, H. and Van der Westhuizen, C. (2008). *Economic growth, poverty and inequality in South Africa: The first decade of democracy*. Paper presented at the DPRU Conference, Muldersdrift.

Bhorat, H., Van der Westhuizen, C. and Goga, S. (2007). *Welfare shifts in the post-apartheid South Africa: A comprehensive measurement of changes*. DPRU Working Paper 07/128. Cape Town: Development Policy Research Unit, University of Cape Town.

Bhorat, H., Van der Westhuizen, C. and Jacobs, T. (2009). *Income and non-income inequality in post-apartheid South Africa: What are the drivers and possible policy interventions?* DPRU Working Paper 09/138. Cape Town: Development Policy Research Unit, University of Cape Town.

Blundell, R. and Preston, I. (1998). Consumption inequality and income uncertainty. *Quarterly Journal of Economics*. 113(2): 604 – 640.

Boltvinik, J. (2001). *Poverty measurement methods – an overview*. New York: United Nations Development Programme (UNDP). [Online] Available:
http://www.flacso.or.cr/fileadmin/documentos/FLACSO/CLPoverty_Measurement_Methods.pdf

Branson, N. (2009). *Re-weighting the OHS and LFS national household survey data to create a consistent series over time: A cross entropy estimation approach*. SALDRU Working Paper Number 38. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

Brookes, M. and Hinks, T. (2004). The racial employment gap in South Africa. *South African Journal of Economics*. 72(3): 573 – 597.

Browning, M., Crossley, T.F. and Weber, G. (2002). *Asking consumption questions in general purpose surveys*. SEDAP Research Paper No. 77. Hamilton: The Program for Research on Social and Economic Dimensions of an Aging Population (SEDAP).

Burger, R., Van der Berg, S., Van der Walt and Yu, D. (2004). *Geography as destiny: Considering the spatial dimensions of poverty and deprivation in South Africa*. Paper presented at the DPRU/TIPS/Cornell University Conference, Somerset West.

Burger, R., Van der Berg, S. and Von Fintel, D. (2012). *The unintended consequences of education policies on South African particiaption and unemployment*. Stellenbosch Economic Working Papers: 11/2012. Stellenbosch: Stellenbosch University.

Burger, R.P. and Woolard, I. (2005). *The state of the labour market in South Africa after the first decade of democracy*. CSSR Working Paper No. 133. Cape Town: Centre for Social Science Research.

Burger, R.P. and Yu, D. (2006). *Wage trends in post-apartheid South Africa: constructing an earnings series from household survey data*. Stellenbosch Economic Working Papers 04/2006. Stellenbosch: Stellenbosch University.

Callan, T. and Nolan, B. (1991). Concepts of poverty and the poverty line. *Journal of Economic Surveys*. 5(3): 243 – 261.

Carter, M. and May, J. (2001). One kind of poverty – poverty dynamics in post-apartheid South Africa. *World Development*. 29(12): 1987 – 2006.

Casale, D. (2004). *What has the feminization of the labour market 'bought' women in South Africa? Trends in labour force participation, employment and earnings, 1995 – 2001*. DPRU Working Paper 04/84. Cape Town: Development Policy Research Unit.

Casale, D., Muller, C. and Posel, D. (2004). 'Two million net new jobs: A reconsideration of the rise in employment in South Africa, 1995-2003. *South African Journal of Economics*. 72(5): 978 – 1002.

Casale, D. and Posel, D. (2002). The continued feminisation of the labour force in South Africa: An analysis of recent data trends. *South African Journal of Economics*. 70(1): 156 – 184.

Chamber, R. (1988). *Poverty in India: Concepts, research and reality*. IDS Discussion Paper 241. Sussex: Institute of Development Studies, University of Sussex.

Cloutier, N.R. (1988). Pareto extrapolation using grouped income data. *Journal of Regional Science*. 28(3): 415-419.

Corti, L. (1993). Using diaries in social research. *Social Research Update*. March 1993. Guildford: University of Surrey.

Cowell, F.A. (1995). *Measuring inequality*. 2nd edition. London: Prentice Hall / Harvester Wheatsheaf.

Datafirst. *Calibrating Stats SA's national household survey weights to a consistent series over time*. Available: http://www.datafirst.uct.ac.za/wiki/index.php?title=File:Calibrating_StatsSA's_Survey_Weights_lw.pdf

Davern, M., Rodin, H., Beebe, T.J. and Thiede Call, K. (2005). The effect of income question design in health surveys on family income, poverty and eligibility estimates. *Health Services Research*. 40(5): 1534-1552.

Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Baltimore: The John Hopkins University Press.

Deaton, A. (2001). Counting the world's poor: Problems and possible solutions. *World Bank Research Observer*. 16(2): 125 – 147.

Deaton, A. (2003). How to monitor poverty for the millennium development goals. *Journal of Human Development*. 4(3): 353-378.

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *The Review of Economics and Statistics*. 87(1): 1-19.

Deaton, A. (2010). Measuring poverty in a growing world (or measuring growth in a poor world). In Anand, S., Segal, P. and Stiglitz, J.E. (ed.), *Debates on the measurement of global poverty*. 1st edition. New York: Oxford University Press: 187-224.

Deaton, A. and Grosh, M. (2000). Consumption. In Grosh, M. and Glewwe, P. (ed.), *Designing household survey questionnaire for developing countries: Lessons from 15 years of the living standards measurement study – Volume One*. Washington: The World Bank: 91 – 133.

Deaton, A. and Kozel, V. (2005). Data and dogma: The great Indian poverty debate. *The World Bank Research Observer*. 20(2): 177-199.

Deaton, A. and Paxson, C. (1997). *Poverty among children and the elderly in developing countries*. Princeton: Research Program in Development Studies, Princeton University.

Devey, R, Skinner, C and Valodia, I. (2006). Definitions, data and the informal economy in South Africa: A critical analysis. In: Padayachee, V (ed.), *The development decade? Economic and social change in South Africa, 1994 – 2004*. Cape Town: Human Sciences Research Council Press: 302-323.

Dias, R. and Posel, D. (2006). *Unemployment, education and skills constraints in post-apartheid South Africa*. DPRU Working Paper 07/120. Cape Town: Development Policy Research Unit, University of Cape Town.

Duclo, J. and Araar, A. (2006). *Poverty and equity: Measurement, policy and estimation with DAD*. 1st edition. Ottawa: Springer.

Edgren, G. (2005). *How different is the SA labour market? International perspectives and parallels*. Pretoria: Human Sciences Research Council, 2005.

Essop, H. and Yu, D. (2008a). *The South African informal sector (1997 – 2006).* Stellenbosch Economic Working Papers: 03/2008. Stellenbosch: Stellenbosch University.

Essop, H. and Yu, D. (2008b). *Alternative definitions of informal sector employment in South Africa*. Stellenbosch Economic Working Papers: 21/2008. Stellenbosch: Stellenbosch University.

Fields, G.S. (1980). *Poverty, inequality and development*. 1st edition. Cambridge: Cambridge University Press.

Fields, G.S. (1989). *A compendium of data on inequality and poverty for the developing world*. Unpublished report. New York: Cornell University.

Finn, A., Leibbrandt, M. and Woolard, I. (2009). *Income and expenditure inequality: Analysis of the NIDS wave 1 dataset*. Discussion Paper No. 5. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

Foster, J.E., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*. 52(3): 761 – 766.

Gelb, S. (2004). *Inequality in South Africa: Nature, causes and responses*. Paper presented at the DPRU/TIPS Conference, Somerset West.

Golan, A., Judge, G. and Miller, D. (1996). *Maximum entropy economics, robust estimation with limited data*. West Sussex: John Wiley and Sons Ltd.

Govender, P., Kambaran, N., Patchett, N., Ruddle, A., Torr, G. and Van Zyl, N. (2006). *Poverty and inequality in South African and the world*. Cape Town: Actuarial Society of South Africa.

Guenard, C. and Mesple-Somps, S. (2010). Measuring inequalities: Do household surveys paint a realistic picture? *Review of Income and Wealth*. 56(3): 519-538.

Gustavsson, M. (2004). *Trends in the transitory variance of earnings: Evidence from Sweden 1960-1990 and a comparison with the United States*. Uppsala University Economics Working Paper 2004: 11. Uppsala: Uppsala University.

Haughton, J. and Khandker, S.R. (2009). *Handbook on poverty and inequality*. Washington: The World Bank.

Haussmans, R. (2007). *Measurement of employment, unemployment and underemployment – current international standards and issues in their application*. [Online] Available: http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/publication/wcms_088394.pdf

Havinga, I., Kamanou, G. and Viet, V.Q. (2010). A note on the (mis)use of national accounts for estimation of household final consumption expenditures for poverty measures. In Anand, S., Segal, P. and Stiglitz, J.E. (ed.), *Debates on the measurement of global poverty*. 1st edition. New York: Oxford University Press: 238-245.

Hlekiso, T. and Mahlo, N. (2009). *An overview of the demand and supply of skills in the South African labour market*. Paper presented at the ESSA Conference, Port Elizabeth.

Hodge, D. (2009). Growth, employment and unemployment in South Africa. *South African Journal of Economics*. 77(4): 488 – 504.

Hoogeveen, J.G. and Özler, B. (2006). Poverty and inequality in post-apartheid South Africa: 1995 – 2000. In Bhorat, H. and Kanbur, R. (ed.), *Poverty and policy in post-apartheid South Africa*. Cape Town: Human Sciences Research Council: 59 – 94.

Karshenas, M. (2003). Global poverty: National accounts based versus survey based estimates. *Development and Change*. 34(4): 683 – 712.

Kingdon, G. and Knight, J. (2004). Unemployment in South Africa: The nature of the beast. *World Development*. 32(3): 391 – 408.

Kingdon, G. and Knight, J. (2007). Unemployment in South Africa, 1995-2003: Causes, Problems and Policies. *Journal of African Economies*. 16(5): 813 – 848.

Klasen, S. (1997) Poverty and inequality in South Africa: An analysis of the 1993 SALDRU survey. *Social Indicators Research*. 41: 51 – 94.

Klasen, S. (2000). Measuring poverty and deprivation in South Africa. *Review of Income and Wealth*. 46(1): 33 – 58.
Klasen, S. and Woolard, I. (1999). Levels, trends and consistency of employment and unemployment figures in South Africa. *Development Southern Africa*. 16(1): 3 – 35.

Lacerda, M., Ardington, C. and Leibbrandt, M. (2008). *Sequential regression multiple imputation for incomplete multivariate data using Markov chain Monte Carlo*. SALDRU paper series number 13. Cape Town: Southern African Labour and Development Research Unit.

Leibbrandt, M., Levinsohn, J. and McCrary, J. (2005). *Incomes in South Africa since the fall of apartheid*. NBER working paper 11384. Cambridge, Massachusetts Avenue: National Bureau for Economic Research.

Leibbrandt, M., Poswell, L., Naidoo, P. and Welch, M. (2006). Measuring recent changes in South African inequality and poverty using 1996 and 2001 Census data. In Bhorat, H. and Kanbur, R. (ed.), *Poverty and policy in post-apartheid South Africa*. Cape Town: Human Sciences Research Council: 95 – 142.

Leibbrandt, M., Woolard, I., Finn, A. and Argent, J. (2010). *Trends in South African income distribution and poverty since the fall of apartheid*. OECD Social, Employment and Migration Working Papers No. 101. Paris: Organisation for Economic Co-operation and Development

Litchfield, J.A. (1999). *Inequality: Methods and tools*.
Available: http://siteresources.worldbank.org/INTPGI/Resources/Inequality/litchfie.pdf

Mahadea, D. (2003). Employment and growth in South Africa: Hope or despair? *South African Journal of Economics*. 71(1): 21 – 48.

Malherbe, J.E. (2007). *An analysis of income and poverty in South Africa*. Unpublished Master thesis. Stellenbosch: Stellenbosch University.

May, J., Carter, M. and Posel, D. (1995). *The composition and persistence of poverty in rural South Africa: An entitlements approach*. Land and Agricultural Policy Centre Policy Paper No. 15. Johannesburg: Land and Agricultural Policy Centre.

McKay, A. (2000). Should the survey measure total household income? In Grosh, M. and Glewwe, P. (ed.), *Designing household survey questionnaire for developing countries: Lessons from 15 years of the living standards measurement study – Volume Two*. Washington: The World Bank: 83 – 104.

Meth, C. (2006a). *What was the poverty headcount in 2004 and how does it compare to recent estimates by van der Berg et al?* SALDRU Working Paper Series 06/01. Cape Town: Southern African Labour and Development Research Unit, University of Cape Town.

Meth, C. (2006b). *Income poverty in 2004: A second engagement with the recent Van der Berg et al figures*. Working Paper No. 47. Durban: School of Development Studies, University of KwaZulu-Natal.

Meth, C. and Dias, R. (2004). Increases in poverty in South Africa, 1999 – 2002. *Development Southern Africa*. 21(1): 59 – 85.

Mlatsheni, C. and Rospabé, S. (2002). *Why is youth unemployment so high and unequally spread in South Africa*. DPRU Working Paper 02/65. Cape Town: Development Policy Research Unit, University of Cape Town.

Mohr, P. (2007). *Economic Indicators*. 3rd edition. Pretoria: University of South Africa.

Moloi, M. and Strauss, J. (2005). *The SACMEQ II Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education*. Harare: Southern Africa Consortium for Monitoring Educational Quality (SACMEQ).

National Association of Automobile Manufacturers of South Africa (NAAMSA).
[Online] Available: http://www.naamsa.co.za/

National Treasury. *Budget Review* (various issues). Pretoria: Government Printers.

National Treasury (1996). *Growth, employment and redistribution: a macroeconomic strategy*. Pretoria: Government Printers.

National Treasury (2007). *Accelerated and shared growth initiative for South Africa: a summary*. Pretoria: Government Printers.

Oosthuizen, M. (2006). *The post-apartheid labour market: 1995-2004*. DPRU Working Paper 06/103. Cape Town: Development Policy Research Unit.

Organization for Economic Co-ordination and Development (OECD) (2008). *What are equivalence scales?*
[Online] Available: http://www.oecd.org/els/social

Özler, B. (2007). Not separate, not equal, poverty and inequality in post-apartheid South Africa. *Economic Development and Cultural Change*. 55(3): 487 – 529.

Parker, R.N. and Fenwick, R. (1983). The Pareto curve and its utility for open-ended income distributions in survey research. *Social Forces*. 61(3): 872-885.

Pauw, K. and Mncube, L. (2007). *The impact of growth and redistribution on poverty and inequality in South Africa*. DPRU Working Paper 07/126. Cape Town: Development Policy Research Unit, University of Cape Town.

Pauw, K., Oosthuizen, M. and Van der Westhuizen, C. (2006). *Graduate unemployment in the face of skills shortages: A labour market paradox*. DPRU Working Paper 06/114. Cape Town: Development Policy Research Unit.

Posel, D. and Casale D. (2005). *Who replies in brackets and what are the implications for earnings estimates? An analysis of earnings data from South Africa*. Paper presented at the ESSA Conference, Durban.

Posel, D. and Rogan, M. (2009). *Gendered trends in poverty in post-apartheid period: 1997 – 2006*. Paper delivered at the ESSA Conference, Port Elizabeth.

Poswell, L. (2002). *The post-apartheid South African labour market: A status report*. Cape Town: Development Policy Research Unit, University of Cape Town.

Raghunathan, T.E., Lepkowski, J.M., van Hoewyk, J. and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*. 27(1): 85-95.

Ravallion, M. (1992). *Poverty comparisons: A guide to concepts and methods*. Living Standards Measurement Study Working Paper No. 88. Washington: The World Bank.

Ravallion, M. (1998). *Poverty lines in theory and practice*. Living Standards Measurement Study Working Paper No. 133. Washington: The World Bank.

Ravallion, M. (2000). Should poverty measures be anchored to the national accounts? *Economic and Political Weekly*: August 26 – September 2: 2345-2352.

Ravallion, M. (2003). Measuring aggregate welfare in developing countries: How well do national accounts and survey agree? *Review of Economics and Statistics*. 85(3): 645 – 652.

Ravallion, M. (2004). *Pro-poor growth: A primer*. World Bank Policy Research Working Paper 3242. Washington: The World Bank.

Ravallion, M. (2010). The debate on globalization, poverty and inequality: Why measurement matters. In Anand, S., Segal, P. and Stiglitz, J.E. (ed.), *Debates on the measurement of global poverty*. 1st edition. New York: Oxford University Press: 25 – 41.

Riphahn, R.R. and Serfling, O. (2004). *Item non-response on income and wealth questions*. Unpublished manuscript. Basel: University of Basel, Switzerland.

Rogan, M. (2010). *Poverty and headship in post-apartheid South Africa*. Paper delivered at the DPRU/TIPS Conference, Johannesburg.

Samson, M., Babson, O., Haarman, C., Haarman, D., Khathi, G., MacQuene, K. and Van Niekerk, I. (2001). *Impact of social security system on poverty in South Africa*. EPRI Research Paper No. 19. Claremont: Economic Research Policy Institute.

Seekings, J. (2006). *Facts, myths, and controversies: The measurement and analysis of poverty and inequality after apartheid*. Paper presented at the DPRU/TIPS Conference, Johannesburg.

Seiver, D.A. (1979). A note of the measurement of income inequality with income data. *The Review of Income and Wealth*. 25(2): 229-234.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrics*. 44: 219-231.

Sen, A. (1987). *Commodities and capabilities*. 1st edition. Amsterdam: North-Holland.

Shepherd, D. (2008). *Post-apartheid trends in gender discrimination in South Africa: Analysis through decomposition techniques*. Paper presented at the DPRU Conference, Muldersdrift.

Simkins, C. (2004). *What happened to the distribution of income in South Africa between 1995 and 2001?* Unpublished paper. University of the Witwatersrand.
Available: http://www.transformationaudit.org.za/research-database/academic-1/labour-market/2000-2006/smikins/
South African Government. (2011). *The New Growth Path.*
[Online] Available: http://www.info.gov.za/view/DownloadFileAction?id=135748

South African Petroleum Industry Association (SAPIA).
[Online] Available: www.sapia.co.za/

Southern Africa Labour Development Research Unit (1994). *Project for Statistics on Living Standards and Development: Metadata*. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

Southern Africa Labour Development Research Unit (2009). *National Income Dynamics Study Wave 1: User document*. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

Statistics South Africa (1996). *October Household Survey 1996: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (1997a). *October Household Survey 1997: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (1997b). *Income and Expenditure Survey 1995: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (1997c). *Earnings and spending in South Africa: Selected findings of the 1995 Income and Expenditure Survey*. Pretoria: Statistics South Africa.

Statistics South Africa (1998a). *October Household Survey 1998: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (1998b). *Census 1996: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (1999). *October Household Survey 1999: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2000a). *Labour Force Survey 2001:1 – Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2000b). *Labour Force Survey 2000 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2001a). *Labour Force Survey 2001 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2001b). *Labour Force Survey 2001 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2001c). *Discussion paper on comparative labour statistics: Survey of Employment and Earnings in selected industries*. Pretoria: Statistics South Africa.

Statistics South Africa (2002a). *Labour Force Survey 2002 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2002b). *Labour Force Survey 2002 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2002c). *Income and Expenditure Survey 2000: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2002d). *Income and expenditure of households 2000 South Africa*. Statistical release P0111. Pretoria: Statistics South Africa.

Statistics South Africa (2002e). *Earnings and spending in South Africa: Selected findings and comparison from the Income and Expenditure Surveys of October 1995 and October 2000*. Pretoria: Government Printer.

Statistics South Africa (2003a). *Labour Force Survey 2003 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2003b). *Labour Force Survey 2003 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2003c). *Census 1996: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2003d). *General Household Survey 2002: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2003e). *Survey of Employment and Earnings: September 2003 and December 2003*. Pretoria: Statistics South Africa.

Statistics South Africa (2004a). *Labour Force Survey 2004 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2004b). *Labour Force Survey 2004 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2004c). *General Household Survey 2003: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2005a). *Labour Force Survey 2005 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2005b). *Labour Force Survey 2005 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2005c). *General Household Survey 2004: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2006a). *Labour Force Survey 2006 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2006b). *Labour Force Survey 2006 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2006c). *General Household Survey 2005: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2006d). *Income and Expenditure Survey 2005/2006: Report on progress made*. Pretoria: Statistics South Africa.

Statistics South Africa (2006e). *Quarterly Employment Statistics: June 2006*. Pretoria: Statistics South Africa.

Statistics South Africa (2007a). *Labour Force Survey 2007 March: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2007b). *Labour Force Survey 2007 September: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2007c). *General Household Survey 2006: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2008a). *Quarterly Labour Force Survey: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2008b). *Guide to Quarterly Labour Force Survey*. Report number: 02-11-01. Pretoria: Statistics South Africa.

Statistics South Africa (2008c). *Report on the response by Statistics South Africa to recommendations of the International Monetary Fund on improvements to the Labour Force Survey*. Pretoria: Statistics South Africa.
[Online] Available: http://www.statssa.gov.za/qlfs/docs/improvements_to_lfs.pdf

Statistics South Africa (2008d). *Income and expenditure of households 2005/2006: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2008e). *Income and expenditure of households 2005/2006: Analysis of results*. Pretoria: Statistics South Africa.

Statistics South Africa (2008f). *Community Survey 2007: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2008g). *General Household Survey 2007: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2009). *General Household Survey 2008: Metadata*. Pretoria: Statistics South Africa.

Statistics South Africa (2010). *General Household Survey 2009: Metadata*. Pretoria: Statistics South Africa.
Streak, J., Yu, D. & Van der Berg, S. (2009). How invariant is South African child poverty to the choice of equivalence scale or poverty measure? *Social Indicators Research*. 94(2): 183-201.

Sudman, S. and Ferber, R. (1971). Experiments in obtaining consumer expenditures by diary methods. *Journal of American Statistical Association*, 66(336): 725-735.
Todaro, M.P. (2009). *Economic development*. 10th edition. Boston: Pearson Addison Wesley.

Turek, J., Denmead, G. and Sinclair-James, B. (2005). *Poverty estimates in the ACS and other income surveys: What is the impact of methodology?* Washington: Assistant Secretary for Planning and Evaluation.

United Nations Development Programme (UNDP) (2003). *South Africa Human Development Report 2003: The challenge of sustainable development*. Pretoria: United Nations Development Programme of South Africa.

United Nations Statistics Division – Classifications Registry.
[Online] Available: http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5

Van der Berg, S. (2010). *Current poverty and income distribution in the context of South African history*. Stellenbosch Economic Working Papers: 22/2010. Stellenbosch: Stellenbosch University.

Van der Berg, Burger, R., Burger, R.P., Louw, M. and Yu, D. (2005). *Trends in poverty and inequality since political transition*. Stellenbosch Economic Working Papers: 01/2005. Stellenbosch: Stellenbosch University.

Van der Berg, S., Burger, R., Burger, R.P., Louw, M. and Yu, D. (2009). *A series of national account-consistent estimates of poverty and inequality in South Africa*. Stellenbosch Economic Working Papers: 09/2007. Stellenbosch: Stellenbosch University.

Van der Berg, S. and Louw, M. (2004). Changing patterns of South African income distribution: Towards time series estimates of distribution and poverty. *South African Journal of Economics*. 72(3): 546 – 572.

Van der Berg, S., Louw, M. and Du Toit, L. (2008). *Poverty trends since the transition: What we know.* Stellenbosch: Stellenbosch University.

Van der Berg, S., Louw, M. and Yu, D. (2008). Post-transition poverty trends based on an alternative data source. *South African Journal of Economics*. 76(1): 58 – 76.

Van der Westhuizen, C., Goga, S. and Oosthuizen, M. (2006). *Women in the South African labour market: 1995 – 2005*. DPRU Working Paper 06/118. Cape Town: Development Policy Research Unit.

Vermaak, C. (2005). *Trends in income distribution, inequality and poverty in South Africa, 1995 to 2003*. Paper presented at the ESSA Conference, Durban.

Vermaark, C. (2008). *The impact of multiple imputation of coarsened data on estimates of the working poor in South Africa*. Paper presented at the DPRU Conference, Muldersdrift.

Von Fintel, D. (2006). *Earnings bracket obstacles in household surveys – how sharp are the tools in the shed?* Stellenbosch Economic Working Papers: 08/06. Stellenbosch: Stellenbosch University.

Von Fintel, D. (2007). Dealing with earnings bracket responses in household surveys – How sharp are midpoint imputations. *South African Journal of Economics*. 75(2): 293-312.

Wenzel, P. and Stewart, R. (2007). *(Un)sustainable growth, (un)employment and poverty. A pro-poor analysis of South Africa's development performance*. Paper presented at the ESSA Conference, 2007.

Whiteford, A. and McGrath, M. (1994). *The distribution of income in South Africa*. 1st edition. Pretoria: Human Sciences Research Council.

Whiteford, A. and Van Seventer, D.E. (2000). South Africa's changing income distribution in the 1990s. *Studies in Economics and Econometrics*. 24(3): 7 – 30.

Wiseman, V., Conteh, L. and Matovu, F. (2005). Using diaries to collect data in resource-poor settings: questions on design and implementation. *Health Policy and Planning*. 20(6): 393 – 404.

Wittenberg, M. (2010). An introduction to maximum entropy and minimum cross-entropy estimation using Stata. *The Stata Journal*. 10(3): 315-330.

Woolard, I. (2001). *Income inequality and poverty: Methods of estimation and some policy applications for South Africa*. Unpublished PhD thesis. Cape Town: University of Cape Town.

Woolard, I. and Leibbrandt, M. (1999). *Measuring poverty in South Africa*. DPRU Working Paper No. 99/33. Cape Town: Development Policy Research Unit, University of Cape Town.

Woolard, I. and Leibbrandt, M. (2006). *Towards a poverty Line for South Africa: Background note*. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

World Bank (1990). *World Development Report 1990: Poverty*. Washington: The World Bank.

World Bank (2000). *World Development Report 2000/2001: Attacking poverty*. Washington: The World Bank.

Yu, D. (2007). *The Comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys*. Stellenbosch Economic Working Papers: 17/2007. Stellenbosch: Stellenbosch University.

Yu, D. (2008). *The comparability of Income and Expenditure Surveys 1995, 2000 and 2005/2006.* Stellenbosch Economic Working Papers: 11/2008. Stellenbosch: Stellenbosch University.

Yu, D. (2009). *The comparability of Labour Force Survey (LFS) and Quarterly Labour Force Survey (QLFS).* Stellenbosch Economic Working Papers: 08/2009. Stellenbosch: Stellenbosch University.

Yu, D. (2010). *Poverty and inequality trends in South Africa using different survey data*. Stellenbosch Economic Working Papers: 04/2010. Stellenbosch: Stellenbosch University.

Appendix A: Additional information on South African labour market

Table A.1: Labour force participation rates by gender, 1994-2009

	Narrow		Broad	
	Male	**Female**	**Male**	**Female**
OHS 1994	59.2%	39.9%	66.6%	50.6%
OHS 1995	58.2%	38.0%	65.8%	48.5%
OHS 1996	54.2%	36.7%	62.6%	47.0%
OHS 1997	54.9%	36.4%	64.1%	48.7%
OHS 1998	58.4%	40.0%	66.4%	51.1%
OHS 1999	59.4%	44.2%	68.1%	56.1%
LFS 2000a	66.5%	56.6%	73.3%	66.4%
LFS 2000b	66.3%	52.3%	72.1%	62.2%
LFS 2001a	65.9%	53.3%	73.4%	64.8%
LFS 2001b	63.7%	49.4%	71.7%	62.5%
LFS 2002a	65.2%	51.8%	73.5%	64.9%
LFS 2002b	64.2%	49.9%	72.8%	63.7%
LFS 2003a	64.1%	50.5%	72.6%	64.4%
LFS 2003b	62.7%	47.4%	72.6%	63.4%
LFS 2004a	61.9%	47.1%	71.9%	62.8%
LFS 2004b	62.0%	46.2%	72.2%	62.7%
LFS 2005a	62.5%	47.7%	72.5%	63.4%
LFS 2005b	63.7%	49.9%	71.9%	63.8%
LFS 2006a	62.9%	49.6%	72.5%	64.5%
LFS 2006b	63.9%	51.1%	72.0%	64.3%
LFS 2007a	63.0%	49.9%	72.2%	63.7%
LFS 2007b	63.9%	49.7%	72.4%	63.7%
QLFS 2008Q1	65.8%	50.9%	68.8%	55.4%
QLFS 2008Q2	65.5%	50.9%	68.3%	55.0%
QLFS 2008Q3	65.2%	50.5%	68.0%	54.5%
QLFS 2008Q4	64.6%	50.3%	67.9%	54.5%
QLFS 2009Q1	64.8%	50.4%	68.2%	54.8%
QLFS 2009Q2	63.6%	49.2%	67.9%	54.6%
QLFS 2009Q3	61.7%	48.0%	66.4%	53.6%
QLFS 2009Q4	62.2%	47.6%	66.9%	53.6%

Table A.2: Labour force participation rates by race, 1994-2009

	Narrow				**Broad**			
	Black	**Coloured**	**Indian**	**White**	**Black**	**Coloured**	**Indian**	**White**
OHS 1994	44.2%	62.5%	55.6%	65.1%	55.9%	65.8%	57.7%	66.6%
OHS 1995	43.1%	60.1%	56.3%	63.3%	54.3%	65.4%	58.4%	64.6%
OHS 1996	39.7%	59.0%	53.1%	63.7%	51.4%	63.6%	55.6%	64.7%
OHS 1997	40.7%	57.2%	54.4%	62.0%	54.1%	62.1%	56.3%	63.4%
OHS 1998	44.9%	58.2%	54.4%	65.0%	56.5%	64.4%	57.6%	66.4%
OHS 1999	47.2%	62.6%	60.7%	67.7%	59.7%	69.5%	64.3%	69.2%
LFS 2000a	59.2%	67.6%	64.5%	68.7%	68.9%	73.8%	71.1%	71.3%
LFS 2000b	56.7%	64.9%	62.1%	69.4%	66.1%	71.3%	64.3%	70.7%
LFS 2001a	57.2%	65.5%	62.4%	68.3%	68.4%	72.9%	65.7%	70.6%
LFS 2001b	53.2%	64.0%	63.0%	69.0%	65.9%	71.9%	66.6%	70.4%
LFS 2002a	55.5%	67.2%	60.7%	69.2%	68.4%	73.3%	64.4%	71.3%
LFS 2002b	54.1%	64.5%	64.6%	67.8%	67.5%	71.2%	67.9%	69.9%
LFS 2003a	54.1%	66.0%	62.4%	69.5%	67.7%	71.7%	65.3%	71.5%
LFS 2003b	51.4%	63.2%	62.5%	69.5%	67.1%	70.5%	65.2%	71.4%
LFS 2004a	51.0%	63.9%	60.2%	68.0%	66.5%	71.0%	63.4%	70.2%
LFS 2004b	50.6%	61.8%	58.8%	68.8%	66.7%	69.5%	64.3%	70.9%
LFS 2005a	51.8%	63.0%	61.4%	68.5%	67.1%	71.0%	66.3%	70.8%
LFS 2005b	53.9%	63.9%	62.7%	67.8%	66.9%	71.9%	67.0%	69.9%
LFS 2006a	53.4%	63.4%	58.8%	68.3%	67.6%	72.2%	66.4%	71.1%
LFS 2006b	54.9%	64.3%	60.0%	68.0%	67.3%	71.4%	63.8%	71.2%
LFS 2007a	53.9%	63.9%	55.8%	67.7%	67.3%	72.4%	60.6%	70.0%
LFS 2007b	53.9%	61.8%	61.4%	71.2%	67.2%	68.8%	66.9%	72.9%
QLFS 2008Q1	55.3%	65.5%	59.9%	70.5%	59.9%	67.4%	60.8%	70.8%
QLFS 2008Q2	55.5%	64.7%	60.4%	68.9%	59.7%	66.2%	60.8%	69.2%
QLFS 2008Q3	55.0%	64.3%	61.8%	68.8%	59.2%	65.6%	62.1%	69.2%
QLFS 2008Q4	54.7%	64.4%	59.8%	68.3%	59.3%	65.6%	60.3%	68.6%
QLFS 2009Q1	54.4%	66.1%	59.8%	69.9%	59.3%	67.1%	60.8%	70.2%
QLFS 2009Q2	53.2%	64.8%	57.8%	69.2%	59.1%	66.4%	59.5%	69.6%
QLFS 2009Q3	51.6%	64.2%	57.4%	67.2%	57.9%	66.1%	59.2%	68.0%
QLFS 2009Q4	51.5%	64.7%	55.9%	68.4%	58.0%	66.8%	57.9%	69.0%

Table A.3: Narrow labour force participation rates by province, 1994-2009

	WC	**EC**	**NC**	**FS**	**KZN**	**NW**	**GAU**	**MPU**	**LIM**
OHS 1994	64.1%	37.6%	56.4%	52.7%	45.9%	49.8%	60.1%	47.5%	34.3%
OHS 1995	61.7%	35.8%	52.9%	52.3%	44.0%	45.1%	61.9%	43.0%	28.8%
OHS 1996	59.4%	32.9%	47.2%	50.9%	40.7%	39.4%	58.8%	43.5%	27.6%
OHS 1997	59.1%	29.3%	50.2%	48.9%	40.3%	42.6%	60.4%	44.5%	31.3%
OHS 1998	58.4%	35.1%	51.6%	52.7%	45.2%	46.8%	61.6%	51.7%	35.7%
OHS 1999	66.4%	38.5%	54.4%	53.7%	49.4%	46.5%	62.4%	51.6%	38.1%
LFS 2000a	70.8%	56.2%	58.7%	67.2%	59.5%	55.8%	68.1%	59.5%	51.1%
LFS 2000b	68.4%	49.5%	62.8%	62.4%	57.9%	54.5%	70.1%	60.0%	43.2%
LFS 2001a	67.1%	50.2%	61.9%	64.1%	57.4%	54.8%	71.2%	60.1%	43.8%
LFS 2001b	66.3%	46.5%	58.0%	61.6%	52.3%	51.5%	69.3%	54.9%	42.8%
LFS 2002a	68.1%	56.2%	59.6%	64.3%	53.8%	51.9%	67.6%	56.0%	43.7%
LFS 2002b	65.6%	46.7%	58.6%	60.2%	56.5%	53.1%	68.9%	54.9%	40.5%
LFS 2003a	68.3%	48.6%	60.0%	62.2%	53.9%	52.1%	68.3%	55.1%	42.4%
LFS 2003b	67.5%	44.1%	55.3%	58.9%	52.0%	50.4%	67.5%	53.7%	37.7%
LFS 2004a	67.1%	41.7%	55.0%	58.2%	51.2%	51.0%	67.1%	53.5%	37.8%
LFS 2004b	66.0%	45.3%	53.4%	57.0%	49.5%	48.3%	65.9%	54.0%	39.0%
LFS 2005a	65.8%	47.5%	56.4%	59.5%	51.9%	50.6%	65.8%	54.1%	37.4%
LFS 2005b	67.0%	48.4%	54.1%	59.7%	53.5%	52.9%	69.0%	54.1%	39.4%
LFS 2006a	67.8%	52.4%	56.4%	56.7%	51.5%	53.1%	66.3%	55.5%	38.2%
LFS 2006b	67.7%	50.8%	60.3%	55.4%	54.2%	52.7%	69.2%	57.5%	37.9%
LFS 2007a	68.1%	46.4%	60.2%	56.4%	53.3%	54.6%	67.4%	56.8%	39.5%
LFS 2007b	66.9%	46.1%	57.7%	57.3%	52.3%	51.8%	69.6%	57.9%	39.5%
QLFS 2008Q1	67.2%	46.3%	57.5%	60.1%	52.8%	53.4%	73.3%	53.6%	42.5%
QLFS 2008Q2	67.3%	45.6%	55.6%	60.2%	53.4%	53.6%	73.2%	54.8%	41.1%
QLFS 2008Q3	66.6%	45.5%	56.3%	58.6%	52.4%	54.3%	73.0%	54.8%	40.9%
QLFS 2008Q4	66.5%	45.5%	56.8%	57.4%	52.4%	55.1%	72.0%	55.2%	39.5%
QLFS 2009Q1	68.7%	45.8%	53.9%	58.2%	51.0%	55.3%	72.0%	56.3%	40.5%
QLFS 2009Q2	68.1%	46.7%	54.1%	56.5%	47.7%	53.5%	71.5%	55.1%	38.6%
QLFS 2009Q3	68.5%	42.5%	51.1%	56.6%	47.2%	49.8%	69.7%	53.2%	38.5%
QLFS 2009Q4	68.3%	42.8%	54.1%	56.3%	46.4%	48.8%	70.0%	53.7%	39.5%

WC: Western Cape
EC: Eastern Cape
NC: Northern Cape
FS: Free State
KZN: KwaZulu-Natal
NW: North West
GAU: Gauteng
MPU: Mpumalanga
LIM: Limpopo

Table A.4: Broad labour force participation rates by province, 1994-2009

	WC	EC	NC	FS	KZN	NW	GAU	MPU	LIM
OHS 1994	67.7%	49.8%	60.0%	60.2%	54.6%	60.4%	69.2%	55.8%	46.7%
OHS 1995	66.5%	47.3%	59.7%	62.6%	53.1%	56.4%	68.7%	54.9%	39.9%
OHS 1996	63.6%	45.4%	56.0%	58.7%	48.7%	53.8%	68.2%	53.1%	37.9%
OHS 1997	62.8%	42.8%	55.4%	58.2%	54.5%	57.5%	69.1%	52.8%	43.5%
OHS 1998	63.9%	46.0%	60.3%	60.6%	57.4%	58.5%	70.3%	59.5%	45.3%
OHS 1999	70.7%	50.7%	62.8%	62.4%	58.9%	61.3%	73.4%	61.9%	50.5%
LFS 2000a	75.3%	64.9%	67.2%	72.3%	68.1%	68.2%	75.7%	67.4%	63.3%
LFS 2000b	72.1%	60.2%	68.1%	69.5%	65.5%	65.3%	76.4%	66.5%	55.4%
LFS 2001a	72.3%	63.0%	69.9%	70.8%	67.1%	69.4%	77.6%	67.4%	59.4%
LFS 2001b	72.2%	60.0%	68.0%	69.4%	64.1%	66.5%	75.4%	65.3%	59.6%
LFS 2002a	72.4%	66.0%	70.1%	71.0%	65.8%	66.7%	77.0%	67.2%	61.4%
LFS 2002b	70.3%	60.0%	68.4%	69.3%	66.7%	66.8%	77.4%	66.7%	61.6%
LFS 2003a	72.7%	60.9%	70.0%	70.9%	65.6%	65.6%	76.5%	69.3%	62.5%
LFS 2003b	72.5%	59.3%	66.3%	71.5%	65.0%	67.0%	76.8%	68.3%	60.2%
LFS 2004a	72.1%	56.8%	70.6%	69.7%	64.4%	66.5%	75.7%	68.5%	61.4%
LFS 2004b	73.0%	59.1%	66.4%	66.8%	62.7%	66.2%	77.4%	67.7%	61.7%
LFS 2005a	72.1%	61.4%	67.8%	67.8%	65.0%	66.2%	77.2%	67.9%	59.3%
LFS 2005b	72.9%	59.9%	67.7%	66.8%	63.6%	67.7%	78.2%	67.6%	59.2%
LFS 2006a	74.0%	64.6%	67.8%	66.3%	64.5%	66.6%	77.4%	66.5%	60.1%
LFS 2006b	74.4%	59.7%	68.5%	66.1%	64.6%	67.1%	78.5%	67.4%	57.8%
LFS 2007a	74.2%	60.5%	69.9%	66.5%	65.0%	67.8%	77.0%	68.2%	57.4%
LFS 2007b	72.1%	58.4%	66.8%	67.5%	65.2%	66.4%	77.7%	68.2%	59.7%
QLFS 2008Q1	68.6%	52.6%	62.7%	63.3%	55.8%	60.7%	76.3%	58.7%	46.6%
QLFS 2008Q2	68.4%	51.3%	60.2%	63.6%	56.0%	59.5%	75.6%	59.1%	46.3%
QLFS 2008Q3	67.4%	51.3%	60.7%	62.4%	55.7%	58.8%	75.2%	58.6%	46.2%
QLFS 2008Q4	67.6%	52.0%	59.3%	61.8%	55.9%	59.8%	74.1%	59.9%	45.7%
QLFS 2009Q1	69.3%	52.2%	57.2%	62.6%	55.3%	60.2%	73.8%	61.1%	47.2%
QLFS 2009Q2	69.0%	54.3%	58.9%	61.6%	54.7%	58.8%	73.7%	60.1%	45.6%
QLFS 2009Q3	69.6%	50.6%	55.4%	61.4%	54.7%	55.7%	72.3%	58.7%	45.6%
QLFS 2009Q4	69.5%	51.1%	58.3%	61.7%	53.4%	55.5%	73.2%	59.3%	46.4%

WC: Western Cape
EC: Eastern Cape
NC: Northern Cape
FS: Free State
KZN: KwaZulu-Natal
NW: North West
GAU: Gauteng
MPU: Mpumalanga
LIM: Limpopo

Table A.5: Narrow labour force participation rates by educational attainment, 1994-2009

	No schooling	Incomplete primary	Incomplete secondary	Matric	Matric + Cert/Dip	Degree	Above Matric
OHS 1994	42.8%	46.3%	42.7%	64.5%	81.7%	82.0%	81.8%
OHS 1995	40.1%	46.8%	40.0%	61.4%	78.1%	82.0%	79.3%
OHS 1996	34.9%	40.2%	38.6%	59.4%	79.0%	82.9%	80.5%
OHS 1997	36.8%	39.8%	38.3%	61.2%	82.5%	82.4%	82.5%
OHS 1998	41.9%	46.3%	40.3%	65.5%	82.1%	85.0%	83.0%
OHS 1999	43.3%	46.9%	43.3%	66.5%	83.6%	85.2%	84.3%
LFS 2000a	59.1%	62.3%	52.9%	72.9%	84.6%	85.2%	84.8%
LFS 2000b	54.7%	57.6%	50.9%	70.8%	86.0%	88.7%	87.2%
LFS 2001a	55.4%	57.8%	51.1%	71.5%	86.8%	87.4%	87.0%
LFS 2001b	45.9%	51.6%	48.4%	72.3%	84.0%	88.7%	86.0%
LFS 2002a	51.3%	55.1%	50.1%	71.4%	87.1%	88.3%	87.6%
LFS 2002b	47.5%	51.7%	48.6%	72.6%	84.6%	89.9%	86.8%
LFS 2003a	45.9%	54.3%	48.6%	71.1%	88.5%	88.9%	88.7%
LFS 2003b	40.8%	48.7%	46.0%	71.4%	88.2%	89.2%	88.6%
LFS 2004a	40.3%	48.2%	45.8%	69.8%	87.0%	87.6%	87.2%
LFS 2004b	41.4%	46.9%	45.6%	69.8%	86.6%	86.7%	86.7%
LFS 2005a	40.8%	49.9%	46.5%	69.0%	85.7%	88.1%	86.7%
LFS 2005b	44.1%	49.8%	48.5%	71.8%	85.6%	85.4%	85.5%
LFS 2006a	44.5%	50.9%	47.4%	69.8%	84.9%	86.9%	85.7%
LFS 2006b	45.4%	50.9%	48.8%	72.0%	87.3%	87.2%	87.3%
LFS 2007a	41.9%	50.4%	48.1%	70.0%	87.3%	86.1%	86.9%
LFS 2007b	44.0%	49.3%	47.4%	71.3%	87.7%	89.5%	88.5%
QLFS 2008Q1	41.8%	51.2%	48.4%	73.1%	88.9%	90.5%	89.5%
QLFS 2008Q2	43.5%	49.5%	48.2%	73.4%	89.9%	90.0%	89.9%
QLFS 2008Q3	42.0%	48.2%	47.8%	74.1%	89.4%	89.0%	89.2%
QLFS 2008Q4	41.0%	46.6%	47.8%	73.3%	90.1%	88.9%	89.7%
QLFS 2009Q1	40.8%	47.9%	48.1%	70.8%	89.8%	88.8%	89.4%
QLFS 2009Q2	38.2%	45.8%	46.6%	71.1%	88.5%	88.5%	88.5%
QLFS 2009Q3	37.3%	43.0%	44.9%	69.6%	87.6%	88.4%	87.9%
QLFS 2009Q4	37.4%	42.4%	44.7%	71.0%	86.1%	89.8%	87.4%

Table A.6: Broad labour force participation rates by educational attainment, 1994-2009

	No schooling	**Incomplete primary**	**Incomplete secondary**	**Matric**	**Matric + Cert/Dip**	**Degree**	**Above Matric**
OHS 1994	56.0%	59.4%	51.3%	71.1%	83.4%	83.1%	83.3%
OHS 1995	51.8%	59.6%	48.6%	69.5%	80.2%	83.1%	81.1%
OHS 1996	46.7%	53.7%	47.7%	67.5%	80.7%	84.8%	82.2%
OHS 1997	50.9%	54.9%	48.3%	71.5%	84.7%	83.7%	84.4%
OHS 1998	52.7%	58.7%	49.8%	74.8%	84.8%	86.3%	85.3%
OHS 1999	54.4%	59.3%	53.7%	77.5%	87.3%	86.9%	87.2%
LFS 2000a	67.6%	72.7%	61.4%	81.3%	88.6%	88.1%	88.4%
LFS 2000b	63.0%	67.3%	59.2%	79.1%	88.5%	89.6%	89.0%
LFS 2001a	65.4%	69.3%	61.0%	81.5%	90.3%	88.3%	89.5%
LFS 2001b	57.2%	66.2%	59.4%	81.3%	88.3%	90.3%	89.2%
LFS 2002a	61.0%	68.5%	61.5%	82.2%	90.1%	90.2%	90.1%
LFS 2002b	59.1%	65.8%	60.5%	83.0%	88.9%	91.2%	89.9%
LFS 2003a	57.6%	68.2%	60.3%	82.5%	91.6%	90.2%	91.0%
LFS 2003b	53.4%	64.4%	59.9%	84.4%	91.9%	90.8%	91.4%
LFS 2004a	53.3%	64.7%	59.5%	82.2%	90.9%	88.8%	90.0%
LFS 2004b	53.9%	63.4%	60.1%	82.8%	91.2%	87.9%	89.8%
LFS 2005a	53.1%	65.6%	60.2%	82.3%	90.4%	89.5%	90.0%
LFS 2005b	55.3%	63.4%	60.5%	82.8%	88.3%	86.0%	87.4%
LFS 2006a	55.5%	65.8%	60.9%	82.1%	89.3%	88.4%	89.0%
LFS 2006b	54.8%	63.6%	60.2%	83.3%	90.4%	89.1%	89.9%
LFS 2007a	53.0%	64.8%	60.7%	81.4%	90.8%	87.0%	89.4%
LFS 2007b	54.5%	63.5%	59.5%	83.2%	91.3%	89.8%	90.7%
QLFS 2008Q1	45.8%	56.8%	52.6%	76.3%	90.2%	90.9%	90.4%
QLFS 2008Q2	47.1%	55.0%	51.9%	76.4%	91.0%	90.1%	90.6%
QLFS 2008Q3	45.7%	53.5%	51.6%	77.1%	90.2%	89.1%	89.8%
QLFS 2008Q4	45.0%	52.5%	51.8%	76.4%	91.4%	89.3%	90.6%
QLFS 2009Q1	44.7%	53.6%	52.6%	73.9%	90.8%	88.9%	90.1%
QLFS 2009Q2	42.9%	53.1%	51.9%	75.5%	89.7%	89.0%	89.5%
QLFS 2009Q3	41.9%	50.8%	50.6%	74.4%	88.8%	88.8%	88.8%
QLFS 2009Q4	41.8%	50.4%	50.6%	75.7%	88.1%	90.5%	88.9%

Table A.7: Provincial shares of employment, 1994-2009

	WC	EC	NC	FS	KZN	NW	GAU	MPU	LIM
OHS 1994	14.8%	9.9%	2.0%	7.6%	19.3%	7.8%	25.8%	6.3%	6.7%
OHS 1995	14.2%	9.7%	2.2%	7.9%	18.0%	7.9%	27.8%	6.1%	6.1%
OHS 1996	15.5%	9.1%	2.4%	7.7%	17.4%	7.6%	27.7%	7.0%	5.6%
OHS 1997	15.3%	8.2%	2.4%	7.6%	17.7%	7.8%	27.8%	6.4%	6.7%
OHS 1998	14.3%	8.5%	2.4%	7.8%	18.3%	7.9%	27.1%	7.2%	6.6%
OHS 1999	15.1%	9.6%	2.3%	7.1%	18.8%	7.5%	26.2%	6.7%	6.7%
LFS 2000a	13.3%	13.3%	2.1%	7.9%	18.6%	7.1%	22.4%	6.3%	8.9%
LFS 2000b	13.1%	11.7%	2.3%	7.5%	19.2%	7.3%	24.6%	6.7%	7.5%
LFS 2001a	12.8%	11.4%	2.2%	7.4%	19.6%	7.5%	24.7%	6.9%	7.6%
LFS 2001b	14.0%	11.1%	2.2%	7.8%	17.9%	7.7%	25.3%	6.5%	7.6%
LFS 2002a	13.9%	14.0%	2.1%	7.3%	17.2%	7.2%	24.7%	6.4%	7.3%
LFS 2002b	13.9%	11.0%	2.2%	7.2%	18.7%	7.6%	25.7%	6.5%	7.2%
LFS 2003a	14.6%	11.9%	2.1%	7.3%	17.8%	7.4%	25.5%	6.5%	6.9%
LFS 2003b	14.5%	10.6%	2.0%	7.1%	18.4%	7.5%	26.1%	6.8%	6.9%
LFS 2004a	15.3%	9.9%	2.1%	7.4%	17.7%	7.4%	26.4%	6.8%	7.0%
LFS 2004b	14.5%	11.0%	2.0%	6.7%	18.0%	7.2%	26.4%	6.8%	7.6%
LFS 2005a	14.4%	11.7%	1.9%	6.7%	17.8%	7.4%	27.1%	6.4%	6.7%
LFS 2005b	14.0%	11.0%	1.9%	6.5%	17.7%	7.5%	28.0%	6.3%	7.1%
LFS 2006a	14.8%	13.1%	2.0%	6.2%	17.6%	7.1%	26.5%	6.4%	6.3%
LFS 2006b	14.4%	10.6%	2.4%	6.1%	19.0%	6.3%	28.0%	7.1%	6.1%
LFS 2007a	14.6%	11.0%	2.1%	6.3%	18.1%	7.2%	27.2%	6.7%	6.8%
LFS 2007b	13.8%	10.1%	2.3%	6.2%	16.9%	6.3%	30.4%	7.4%	6.6%
QLFS 2008Q1	13.9%	9.7%	2.2%	6.2%	18.8%	6.6%	29.5%	6.6%	6.5%
QLFS 2008Q2	13.8%	9.9%	2.2%	6.1%	19.1%	6.5%	29.6%	6.6%	6.4%
QLFS 2008Q3	13.7%	9.6%	2.2%	6.2%	18.9%	6.3%	29.8%	6.8%	6.5%
QLFS 2008Q4	13.9%	9.8%	2.3%	6.0%	19.0%	6.5%	29.5%	6.7%	6.3%
QLFS 2009Q1	14.4%	9.6%	2.0%	5.9%	18.4%	6.5%	29.5%	6.9%	6.6%
QLFS 2009Q2	14.2%	10.1%	2.1%	5.8%	18.4%	6.4%	29.6%	6.7%	6.8%
QLFS 2009Q3	14.5%	9.8%	2.0%	5.9%	19.1%	6.1%	28.9%	6.8%	7.0%
QLFS 2009Q4	14.6%	9.7%	2.2%	6.1%	18.6%	6.1%	28.9%	6.8%	7.0%

WC: Western Cape
EC: Eastern Cape
NC: Northern Cape
FS: Free State
KZN: KwaZulu-Natal
NW: North West
GAU: Gauteng
MPU: Mpumalanga
LIM: Limpopo

Table A.8: The share of employed of each age cohort, 1994-2009

	15-24 years	25-34 years	35-44 years	45-54 years	55-65 years
OHS 1994	12.4%	33.2%	29.1%	18.2%	7.2%
OHS 1995	11.8%	34.5%	30.1%	16.7%	6.9%
OHS 1996	12.3%	33.3%	30.6%	16.8%	7.1%
OHS 1997	10.9%	33.6%	30.9%	17.7%	6.9%
OHS 1998	11.7%	34.0%	30.5%	17.0%	6.8%
OHS 1999	12.5%	33.9%	29.7%	16.9%	7.0%
LFS 2000a	14.9%	31.2%	28.4%	16.7%	8.8%
LFS 2000b	12.8%	31.9%	27.4%	18.7%	9.2%
LFS 2001a	12.4%	32.0%	27.8%	18.6%	9.3%
LFS 2001b	11.8%	32.7%	28.5%	18.8%	8.3%
LFS 2002a	12.2%	32.3%	28.0%	18.7%	8.8%
LFS 2002b	11.4%	33.4%	28.1%	18.7%	8.5%
LFS 2003a	10.6%	33.7%	28.3%	19.1%	8.4%
LFS 2003b	10.8%	34.3%	27.6%	19.1%	8.3%
LFS 2004a	10.6%	34.1%	27.6%	19.0%	8.6%
LFS 2004b	11.1%	33.9%	26.9%	19.5%	8.6%
LFS 2005a	10.7%	33.6%	27.1%	19.3%	9.3%
LFS 2005b	11.5%	33.8%	26.4%	19.3%	9.0%
LFS 2006a	11.4%	34.0%	25.9%	19.4%	9.3%
LFS 2006b	11.4%	34.0%	26.1%	19.4%	9.0%
LFS 2007a	11.2%	34.4%	26.4%	18.9%	9.0%
LFS 2007b	11.4%	34.2%	26.6%	19.0%	8.8%
QLFS 2008Q1	12.1%	34.0%	26.4%	19.0%	8.5%
QLFS 2008Q2	12.3%	33.8%	26.3%	19.0%	8.5%
QLFS 2008Q3	11.7%	34.1%	26.4%	19.2%	8.5%
QLFS 2008Q4	11.7%	34.1%	26.7%	19.0%	8.5%
QLFS 2009Q1	11.4%	33.4%	27.0%	19.5%	8.8%
QLFS 2009Q2	11.1%	33.4%	27.4%	19.3%	8.7%
QLFS 2009Q3	10.6%	33.5%	27.8%	19.4%	8.6%
QLFS 2009Q4	10.8%	33.4%	27.7%	19.6%	8.6%

Table A.9: The share of employed of each educational attainment category, 1994-2009

	No schooling	**Incomplete primary**	**Incomplete secondary**	**Matric**	**Matric + Cert/Dip**	**Degree**	**Matric or above**
OHS 1994	8.0%	15.9%	41.4%	21.3%	8.4%	4.9%	34.6%
OHS 1995	8.2%	16.3%	39.1%	22.2%	9.4%	4.7%	36.4%
OHS 1996	8.0%	14.7%	38.7%	24.1%	8.7%	5.7%	38.6%
OHS 1997	8.3%	14.1%	40.1%	23.3%	9.3%	4.9%	37.5%
OHS 1998	9.0%	15.9%	37.2%	24.0%	9.3%	4.6%	37.9%
OHS 1999	7.6%	17.1%	37.5%	23.8%	7.4%	6.7%	37.9%
LFS 2000a	8.7%	18.7%	39.6%	20.7%	7.1%	5.2%	32.9%
LFS 2000b	8.2%	18.3%	38.7%	19.8%	8.0%	7.0%	34.8%
LFS 2001a	8.2%	17.5%	38.7%	21.5%	7.9%	6.1%	35.5%
LFS 2001b	7.1%	16.7%	37.3%	24.1%	8.0%	6.7%	38.8%
LFS 2002a	7.9%	16.4%	37.5%	23.5%	8.0%	6.6%	38.1%
LFS 2002b	7.1%	15.5%	37.3%	24.6%	8.5%	7.0%	40.1%
LFS 2003a	6.7%	16.0%	37.3%	24.3%	8.7%	7.1%	40.1%
LFS 2003b	5.9%	14.6%	36.6%	27.0%	8.9%	7.0%	42.9%
LFS 2004a	6.2%	14.7%	36.8%	26.8%	8.6%	7.0%	42.4%
LFS 2004b	6.3%	13.6%	37.6%	27.3%	8.7%	6.5%	42.5%
LFS 2005a	5.5%	13.6%	37.8%	27.5%	8.7%	6.8%	43.1%
LFS 2005b	5.8%	12.9%	38.5%	27.5%	8.9%	6.4%	42.7%
LFS 2006a	5.3%	13.1%	38.0%	28.0%	9.2%	6.4%	43.6%
LFS 2006b	5.2%	12.4%	38.8%	27.9%	9.7%	6.0%	43.6%
LFS 2007a	5.1%	12.2%	39.0%	28.0%	9.7%	6.1%	43.8%
LFS 2007b	5.2%	11.9%	37.6%	27.1%	10.1%	8.0%	45.2%
QLFS 2008Q1	4.3%	11.0%	38.9%	28.9%	10.6%	6.3%	45.9%
QLFS 2008Q2	4.3%	10.8%	38.8%	29.0%	10.5%	6.6%	46.0%
QLFS 2008Q3	4.3%	10.7%	38.6%	29.0%	10.6%	6.8%	46.4%
QLFS 2008Q4	4.1%	10.2%	39.4%	28.5%	11.2%	6.7%	46.3%
QLFS 2009Q1	4.0%	10.4%	38.5%	28.7%	11.6%	6.9%	47.2%
QLFS 2009Q2	3.7%	9.5%	38.9%	29.2%	11.8%	6.9%	47.9%
QLFS 2009Q3	3.6%	9.1%	37.9%	30.0%	12.2%	7.1%	49.3%
QLFS 2009Q4	3.4%	9.2%	38.1%	30.6%	11.6%	7.0%	49.3%

Note: Employed with unspecified educational attainment were excluded.

Table A.10: Proportion of employed in highly-skilled occupations by broad industry category, 1994-2009

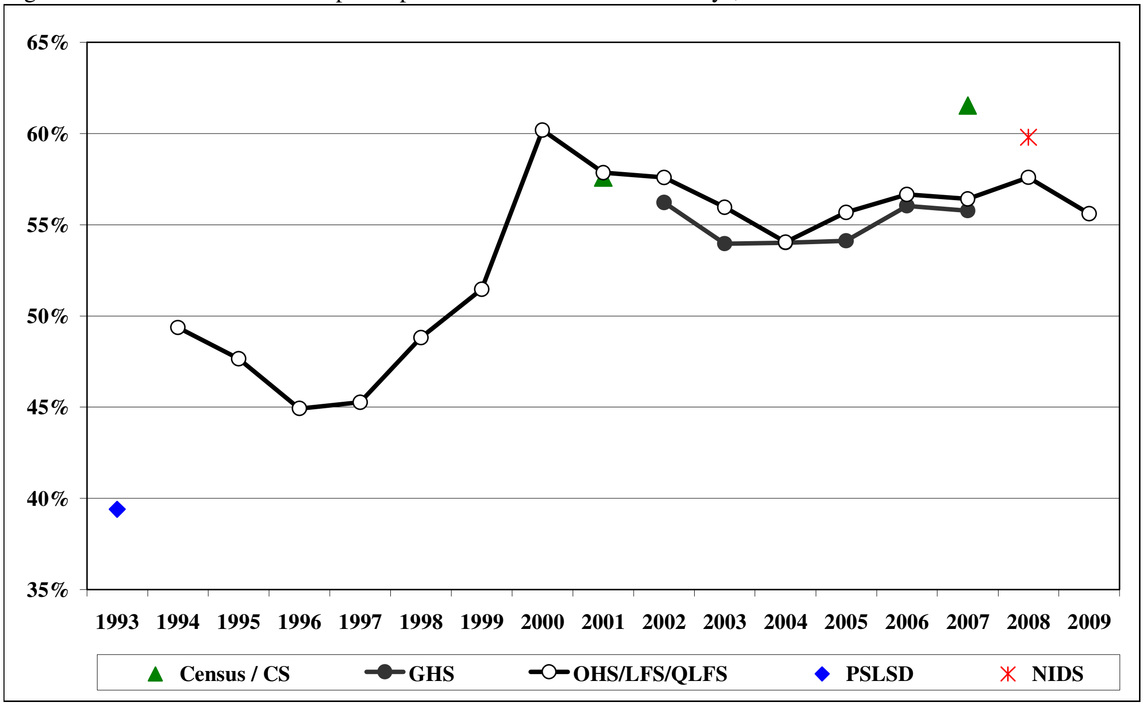
	[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]
OHS 1994	1.2%	7.1%	13.1%	17.3%	10.6%	17.1%	20.7%	44.9%	33.5%
OHS 1995	0.8%	6.6%	11.7%	18.1%	9.5%	16.7%	26.2%	37.6%	45.6%
OHS 1996	3.9%	9.9%	17.6%	22.6%	5.7%	16.0%	25.5%	39.8%	52.9%
OHS 1997	5.0%	18.7%	18.4%	19.1%	8.7%	19.2%	21.5%	39.4%	54.1%
OHS 1998	5.2%	15.1%	17.6%	33.1%	8.5%	16.2%	22.6%	44.4%	50.4%
OHS 1999	3.3%	8.6%	17.2%	19.2%	9.2%	14.6%	23.5%	42.6%	53.0%
LFS 2000a	2.0%	9.4%	17.8%	25.1%	5.6%	12.6%	17.8%	36.8%	51.6%
LFS 2000b	1.8%	9.4%	14.4%	28.8%	5.3%	11.5%	22.8%	40.9%	52.2%
LFS 2001a	1.2%	6.9%	16.4%	20.9%	4.3%	10.0%	25.3%	40.4%	51.4%
LFS 2001b	1.4%	5.9%	16.4%	22.5%	7.2%	12.4%	25.7%	43.3%	52.5%
LFS 2002a	1.8%	6.8%	18.8%	25.1%	6.7%	12.0%	29.2%	44.9%	52.4%
LFS 2002b	1.9%	6.4%	19.0%	21.2%	6.7%	13.3%	29.2%	43.8%	52.0%
LFS 2003a	2.5%	5.2%	18.5%	25.0%	7.4%	12.9%	25.8%	42.7%	50.8%
LFS 2003b	5.2%	7.3%	18.3%	21.0%	8.2%	13.0%	24.3%	44.6%	51.4%
LFS 2004a	4.0%	6.3%	19.4%	30.5%	7.3%	15.0%	25.8%	39.8%	50.7%
LFS 2004b	4.8%	7.1%	17.1%	39.1%	8.8%	14.3%	23.5%	39.2%	49.8%
LFS 2005a	3.9%	7.4%	16.3%	20.9%	6.5%	12.7%	23.5%	39.9%	49.3%
LFS 2005b	4.9%	6.1%	15.6%	22.6%	9.6%	13.5%	20.9%	41.7%	50.7%
LFS 2006a	3.7%	5.6%	17.9%	35.9%	6.3%	14.4%	20.3%	42.6%	50.5%
LFS 2006b	3.9%	7.1%	15.1%	28.1%	8.2%	15.4%	20.6%	39.2%	48.8%
LFS 2007a	4.5%	6.7%	16.6%	25.8%	8.1%	14.7%	21.4%	36.8%	50.1%
LFS 2007b	3.5%	7.9%	19.2%	32.0%	10.8%	17.8%	30.3%	44.8%	55.5%
QLFS 2008Q1	4.9%	9.5%	17.0%	23.5%	10.1%	14.8%	25.0%	39.9%	49.7%
QLFS 2008Q2	5.9%	12.1%	17.5%	31.7%	11.3%	15.3%	23.3%	40.4%	49.6%
QLFS 2008Q3	6.1%	11.1%	18.7%	27.9%	12.0%	16.3%	23.1%	41.0%	50.1%
QLFS 2008Q4	5.5%	13.2%	18.1%	33.6%	12.7%	16.5%	23.1%	40.2%	51.6%
QLFS 2009Q1	4.5%	11.7%	18.2%	22.3%	14.0%	16.1%	21.3%	42.4%	51.9%
QLFS 2009Q2	5.3%	12.1%	18.5%	24.5%	14.9%	14.5%	24.4%	40.7%	49.3%
QLFS 2009Q3	3.1%	10.8%	18.9%	28.9%	13.6%	15.0%	25.4%	39.9%	51.5%
QLFS 2009Q4	5.9%	11.3%	19.5%	27.9%	12.2%	15.1%	25.4%	39.0%	48.4%

Primary: [A]: Agriculture, forestry, fishing and hunting
[B]: Mining and quarrying
Secondary: [C]: Manufacturing
[D]: Electricity, gas and water supply
[E]: Construction
Tertiary: [F]: Wholesale and retail
[G]: Transport, storage and communication
[H]: Financial, insurance and business services
[I]: Community, social and personal services

Table A.11: Narrow unemployment rates by educational attainment, 1994-2009

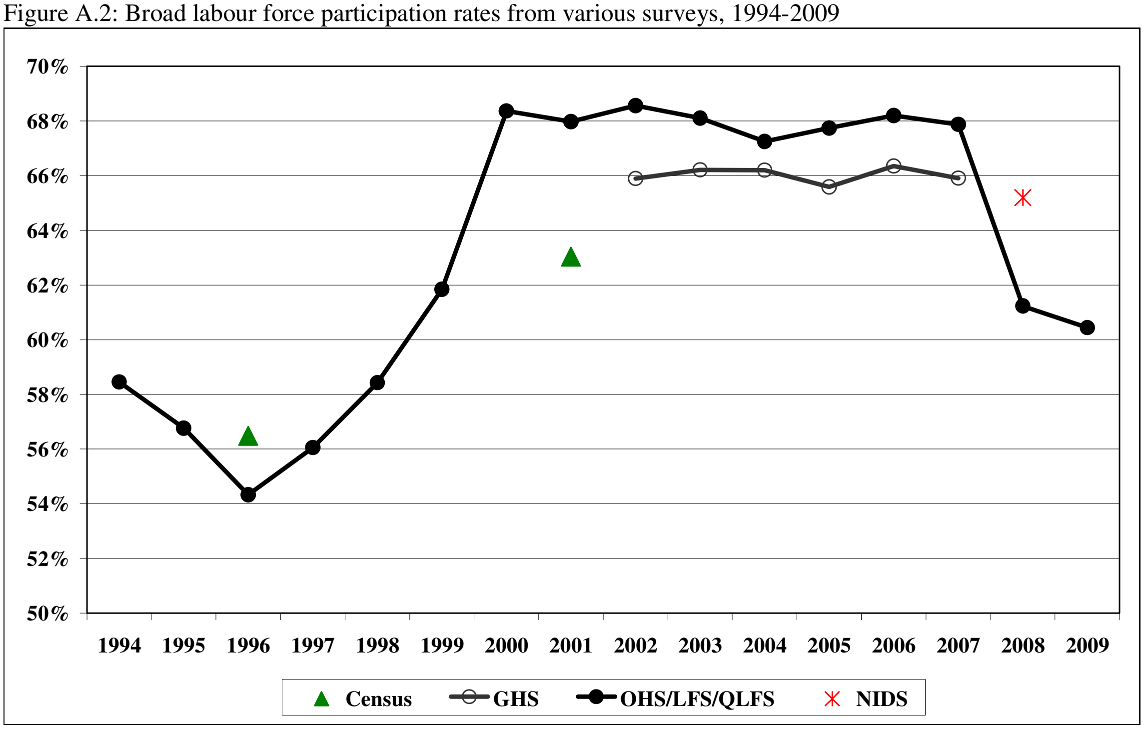
	No schooling	Incomplete primary	Incomplete secondary	Matric	Matric + Cert/Dip	Degree	Above Matric
OHS 1994	20.1%	25.7%	24.3%	17.8%	4.6%	2.3%	12.9%
OHS 1995	15.5%	19.6%	21.3%	17.3%	5.4%	2.5%	12.8%
OHS 1996	19.1%	25.5%	24.0%	17.2%	4.1%	3.2%	12.6%
OHS 1997	18.6%	25.3%	24.6%	21.1%	7.5%	4.0%	16.1%
OHS 1998	20.7%	28.0%	30.3%	24.5%	10.2%	4.3%	19.3%
OHS 1999	17.1%	24.0%	28.5%	24.5%	11.9%	4.7%	19.3%
LFS 2000a	15.8%	24.3%	30.9%	31.4%	16.9%	7.3%	25.6%
LFS 2000b	17.0%	23.8%	30.1%	29.7%	14.5%	4.8%	22.4%
LFS 2001a	15.2%	24.1%	31.1%	31.5%	13.8%	7.0%	24.6%
LFS 2001b	21.6%	28.1%	35.8%	30.8%	14.9%	6.8%	24.5%
LFS 2002a	17.5%	26.7%	36.2%	32.7%	16.7%	6.2%	26.1%
LFS 2002b	19.5%	29.9%	37.3%	32.0%	14.3%	6.1%	25.1%
LFS 2003a	21.2%	28.5%	37.7%	34.3%	14.6%	6.7%	26.9%
LFS 2003b	17.8%	26.6%	34.7%	29.3%	13.5%	3.9%	23.1%
LFS 2004a	14.6%	23.8%	34.5%	31.5%	11.6%	5.5%	24.7%
LFS 2004b	14.8%	25.3%	32.9%	27.7%	9.9%	3.2%	21.5%
LFS 2005a	18.0%	24.4%	33.0%	27.7%	11.5%	3.2%	21.7%
LFS 2005b	18.1%	26.3%	32.6%	28.2%	10.5%	3.8%	22.0%
LFS 2006a	19.9%	22.6%	31.3%	27.1%	12.7%	3.8%	21.6%
LFS 2006b	19.3%	24.0%	31.4%	26.7%	10.3%	3.7%	20.9%
LFS 2007a	17.0%	22.6%	31.9%	26.8%	9.8%	4.3%	20.9%
LFS 2007b	12.6%	22.6%	29.5%	23.2%	10.3%	2.5%	17.4%
QLFS 2008Q1	15.6%	22.6%	29.1%	24.4%	10.5%	5.2%	19.2%
QLFS 2008Q2	16.1%	21.3%	28.6%	24.3%	10.3%	3.3%	18.9%
QLFS 2008Q3	13.8%	21.5%	29.1%	24.4%	10.6%	3.2%	18.9%
QLFS 2008Q4	12.6%	20.8%	27.3%	23.6%	8.7%	2.7%	17.8%
QLFS 2009Q1	14.9%	19.3%	29.5%	25.5%	10.5%	3.5%	19.5%
QLFS 2009Q2	16.1%	23.3%	28.6%	25.7%	10.2%	3.8%	19.7%
QLFS 2009Q3	16.4%	24.1%	30.7%	25.7%	11.4%	4.2%	19.9%
QLFS 2009Q4	17.6%	23.0%	30.0%	25.7%	10.9%	4.5%	20.1%

Figure A.1: Narrow labour force participation rates from various surveys, 1993-2009



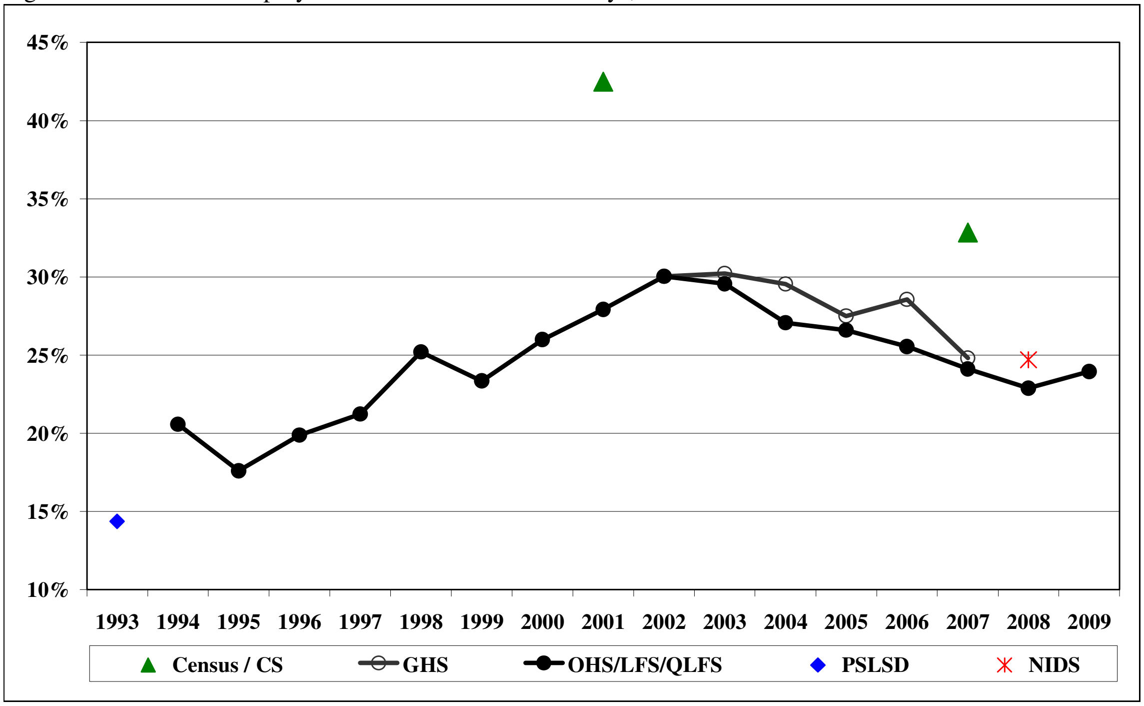
Note: The LFS/QLFS 2000 – 2009 rates are derived by taking the average of the figures from the surveys taking place in the same year, i.e., the LFS 2000a and LFS 2000b rates are averaged to derive the 2000 rate, and so forth.

Figure A.2: Broad labour force participation rates from various surveys, 1994-2009



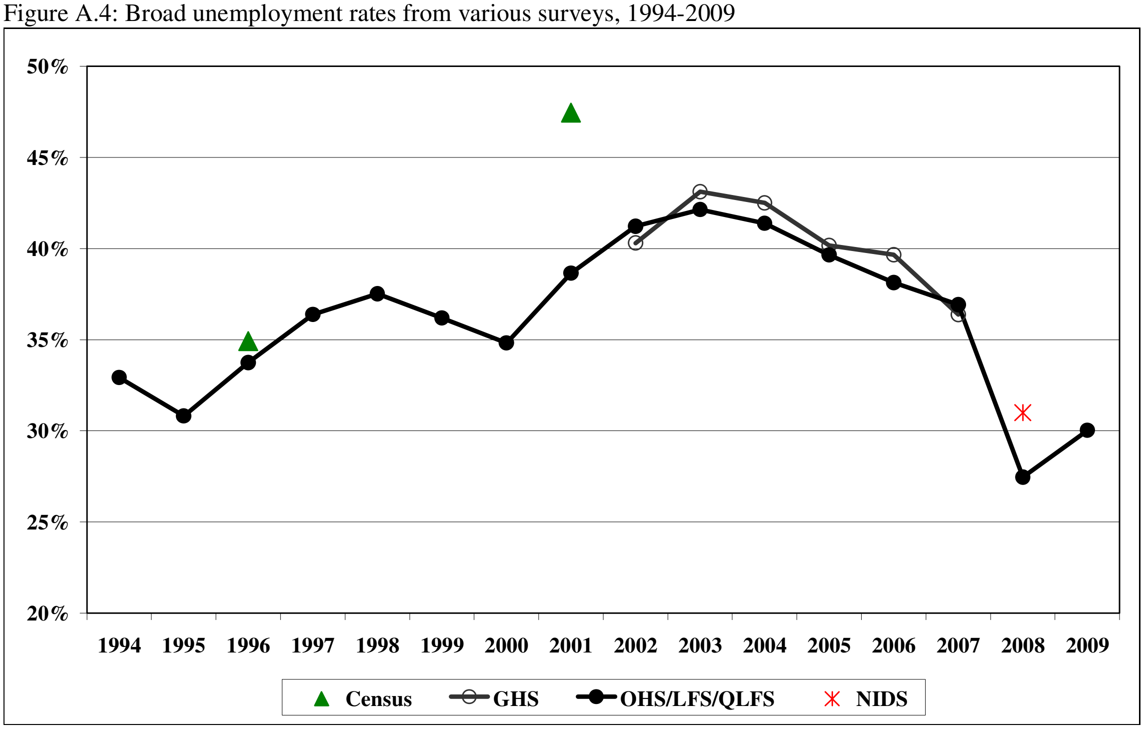
Note: The LFS/QLFS 2000 – 2009 rates are derived by taking the average of the figures from the surveys taking place in the same year, i.e., the LFS 2000a and LFS 2000b rates are averaged to derive the 2000 rate, and so forth.

Figure A.3: Narrow unemployment rates from various surveys, 1994-2009



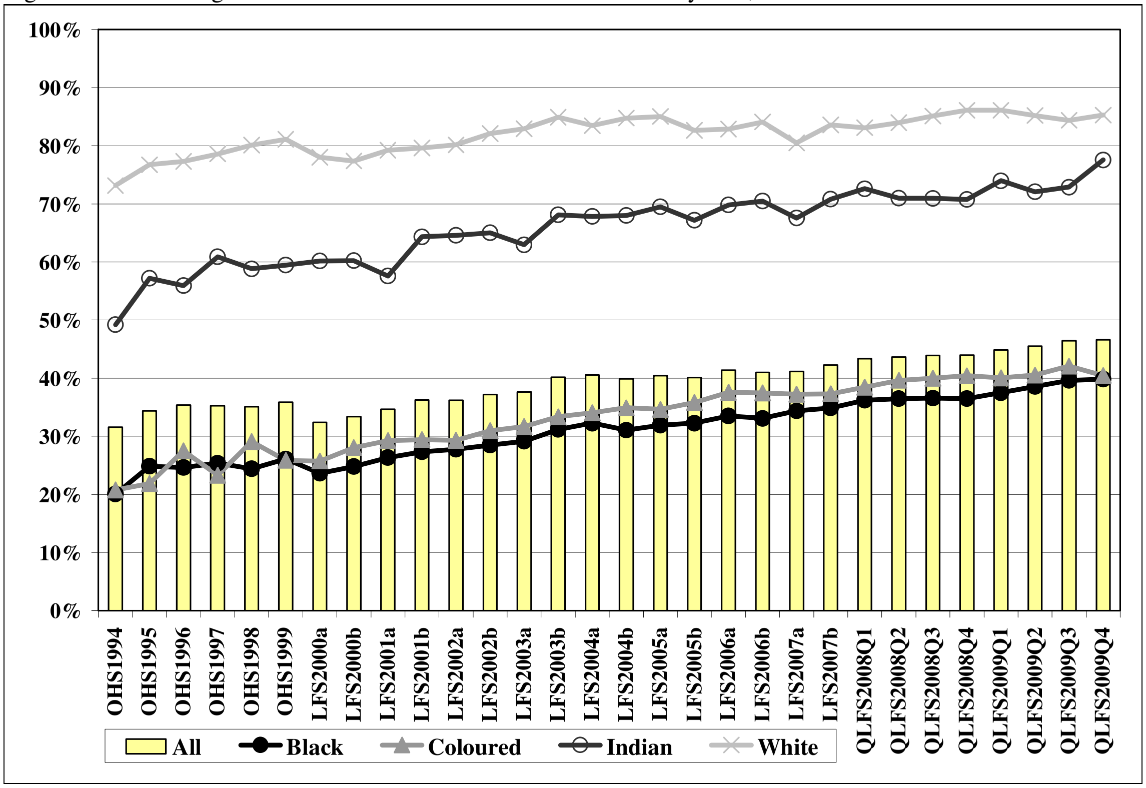
Note: The LFS/QLFS 2000 – 2009 rates are derived by taking the average of the figures from the surveys taking place in the same year, i.e., the LFS 2000a and LFS 2000b rates are averaged to derive the 2000 rate, and so forth.

Figure A.4: Broad unemployment rates from various surveys, 1994-2009



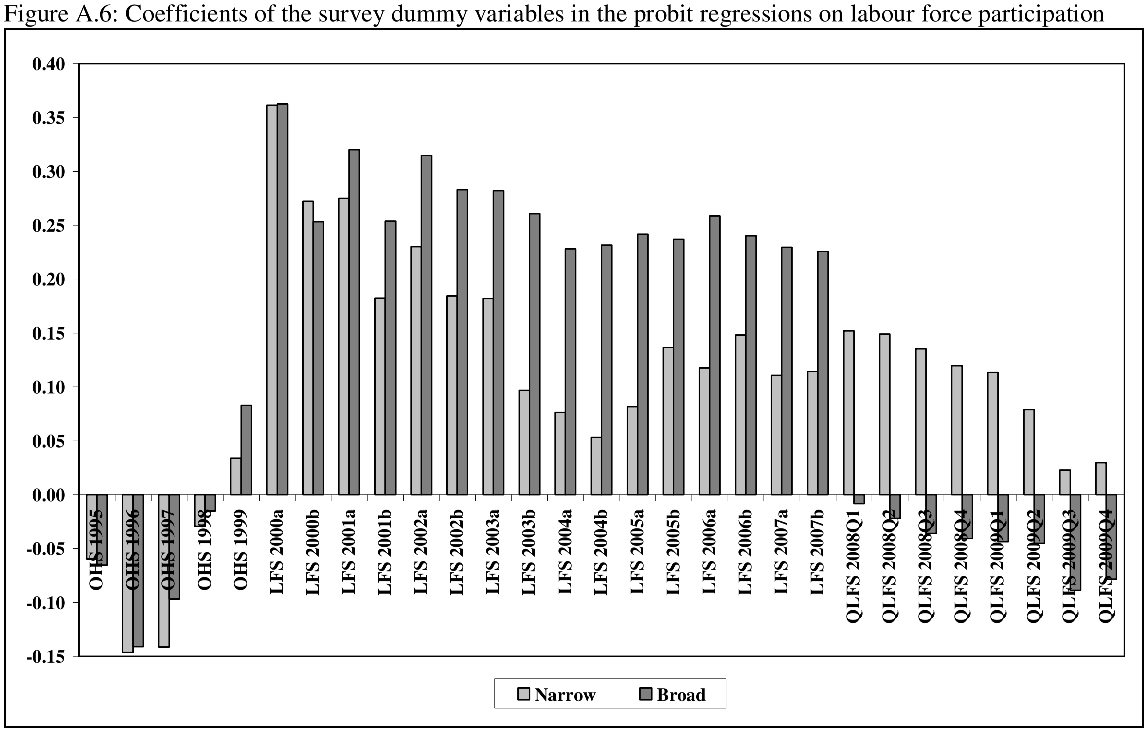
Note: The LFS/QLFS 2000 – 2009 rates are derived by taking the average of the figures from the surveys taking place in the same year, i.e., the LFS 2000a and LFS 2000b rates are averaged to derive the 2000 rate, and so forth.

Figure A.5: Percentage of narrow labour force with at least Matric by race, 1994-2009



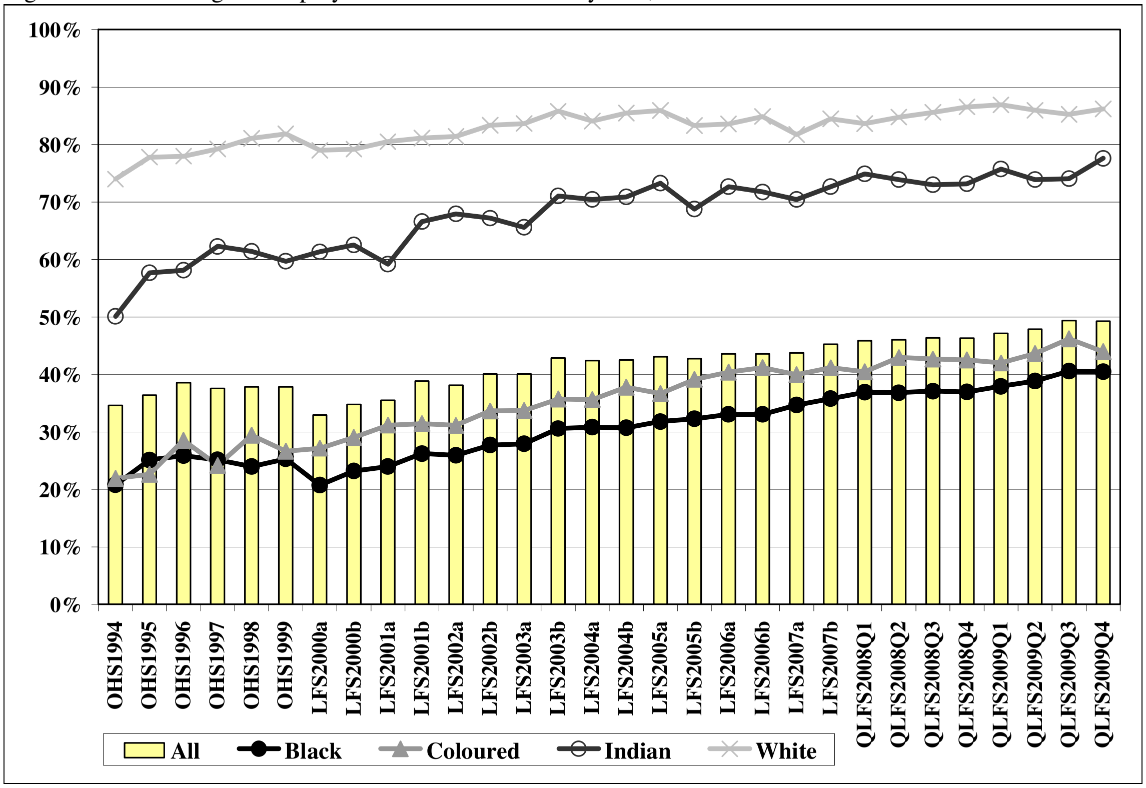
Note: The labour force with unspecified educational attainment was excluded.

Figure A.6: Coefficients of the survey dummy variables in the probit regressions on labour force participation



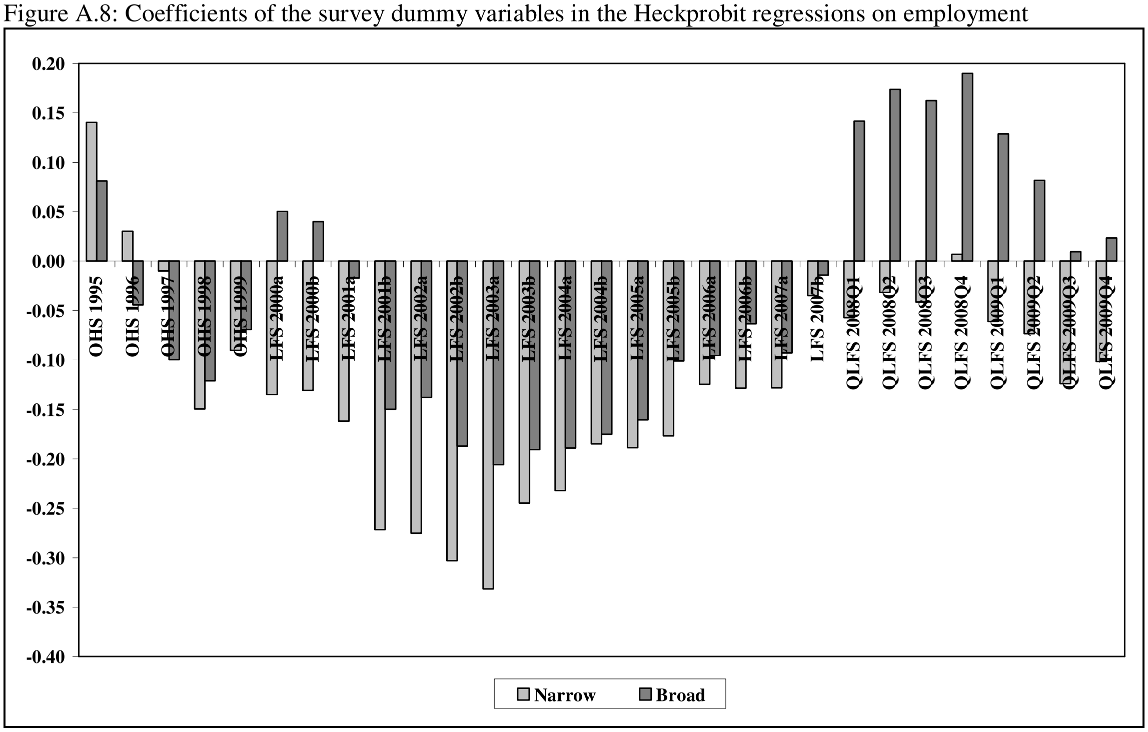
Note: Reference group: OHS 1994

Figure A.7: Percentage of employed with at least Matric by race, 1994-2009



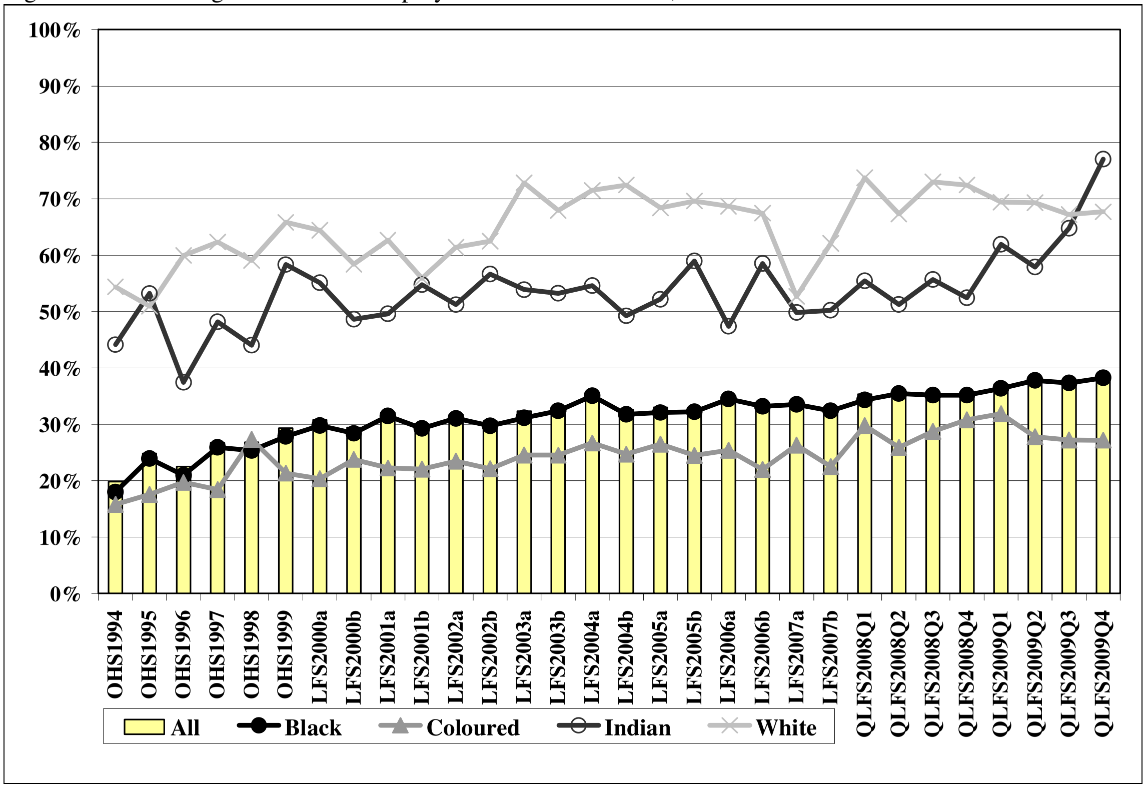
Note: The employed with unspecified educational attainment were excluded.

Figure A.8: Coefficients of the survey dummy variables in the Heckprobit regressions on employment



Note: Reference group: OHS 1994

Figure A.9: Percentage of narrow unemployed with at least Matric, 1994-2009



Note: The unemployed with unspecified educational attainment were excluded.

Appendix B: Stata do-files on labour market status derivation

Table B.1: Stata do-files on the derivation of labour market status in OHS 1996-LFS 2000a using the LFS 2000b-LFS 2007b methodology

```
* LFS 2000a
gen s1 = 0
rename q31_ynotw reason
replace s1 = 1 if reason == 1 | reason == 2 | reason == 8 | reason == 9 | reason == 10 | reason == 11 |
reason == 12
replace s1 = 0 if (q35_accep >= 2 | q36_whnst > 2) & s1 == 2
replace s1 = 0 if q37_lookw  > 1 & q37_bgnbu > 1 & s1 == 2
replace s1 = 1 if reason == 1
gen s2 = 0
replace s2 = 1 if reason == 1 | reason == 2 | reason == 8 | reason == 9 | reason == 10 | reason == 11 |
reason == 12
replace s2 = 0 if (q35_accep >= 2 | q36_whnst > 2) & s2 == 2
replace s2 = 1 if reason == 1
tab s1 if age >= 15 & age <= 65 [w=int(weight)]
tab s2 if age >= 15 & age <= 65[w=int(weight)]

* OHS 1999
gen s1 = 0
rename Q3_37NOT reason
replace s1 = 1 if reason == 2 | reason == 1 | reason == 8 | reason == 9 | reason == 10 | reason == 11
replace s1 = 0 if (Q3_29ACC >= 2 | Q3_30STA > 2)
replace s1 = 0 if (Q3_32AWH  ==  1)
replace s1 = 1 if reason == 2
gen s2 = 0
replace s2 = 1 if reason == 2 | reason == 1 | reason == 8 | reason == 9 | reason == 10 | reason == 11
replace s2 = 0 if (Q3_29ACC >= 2 | Q3_30STA > 2)
replace s2 = 1 if reason == 2
tab s1 if C1_AGE >= 15 & C1_AGE <= 65 [w=int(WGT4)]
tab s2 if C1_AGE >= 15 & C1_AGE <= 65 [w=int(WGT4)]

* OHS 1998
gen s1 = 0
rename q31_ynotw reason
replace s1 = 1 if reason == 2 | reason == 1 | reason == 8 | reason == 9 | reason == 10 | reason == 11
replace s1 = 0 if (q35_accep == 2 | q36_whnst == 5)
replace s1 = 0 if (Q3_34AWH  ==  1)
replace s1 = 1 if reason == 2
gen s2 = 0
replace s2 = 1 if reason == 2 | reason == 1 | reason == 8 | reason == 9 | reason == 10 | reason == 11
replace s2 = 0 if (q35_accep == 2 | q36_whnst == 5)
replace s2 = 1 if reason == 2
tab s1 if age >= 15 & age <= 65 [w=int(weight/10000)]
tab s2 if age >= 15 & age <= 65 [w=int(weight/10000)]

* OHS 1997
gen s1 = 0
rename WOCCDESC reason
replace s1 = 1 if reason == 2 | reason == 1 | reason == 8 | reason == 9 | reason == 10 | reason == 11
replace s1 = 0 if (WACCJOB == 2 | WSTRTSOO == 5)
replace s1 = 0 if (WWHFND1  ==  1)
replace s1 = 1 if reason == 2
gen s2 = 0
replace s2 = 1 if reason == 2 | reason == 1 | reason == 8 | reason == 9 | reason == 10 | reason == 11
replace s2 = 0 if (WACCJOB == 2 | WSTRTSOO == 5)
replace s2 = 1 if reason == 2
tab s1 if PAGE >= 15 & PAGE <= 65 [w=int(PERSWGT)]
tab s2 if PAGE >= 15 & PAGE <= 65 [w=int(PERSWGT)]
```

Table B.1: Continued

```
* OHS 1996
gen s1 = 0
rename RNOTWORK reason
replace s1 = 1 if reason == 3 | reason == 1 | reason == 9 | reason == 2 | reason == 10 | reason == 11
replace s1 = 0 if (EJOBSOFF == 2 | SOONSTAR == 5)
replace s1 = 0 if (FINDWORK  ==  1)
replace s1 = 1 if reason == 3
gen s2 = 0
replace s2 = 1 if reason == 3 | reason == 1 | reason == 9 | reason == 2 | reason == 10 | reason == 11
replace s2 = 0 if (EJOBSOFF == 2 | SOONSTAR == 5)
replace s2 = 1 if reason == 3
tab s1 if PERSONAG >= 15 & PERSONAGE <= 65 [w=int(NEWPWGT)]
tab s2 if PERSONAG >= 15 & PERSONAG <= 65 [w=int(NEWPWGT)]
```

Note: s1 = Narrow unemployed dummy variable (1: unemployed, 0: employed or inactive)
s2 = Broad unemployed dummy variable (1: unemployed, 0: employed or inactive)

Table B.2: Stata do-files on the derivation of labour market status in the LFSs using the revised QLFS methodology

```
* QLFSs
gen sss = 0
replace sss = 1 if q31alookwrk == 1 & q39joboffer == 1
replace sss = 1 if q31bstartbusns == 1 & q39joboffer == 1
replace sss = 1 if q33havejob == 1 & q39joboffer == 1
replace sss = 2 if q38rsnnotseek == 8 & q39joboffer == 1
replace sss = 2 if q38rsnnotseek == 10 & q39joboffer == 1
replace sss = 2 if q38rsnnotseek == 11 & q39joboffer == 1
tab sss if q14age >= 15 & q14age <= 65 [w=int(weight)]

* LFS 2007b
gen sss = 0
replace sss = 1 if Q38LookW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q38BgnBu == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q32YnotW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 6 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 8 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 9 & Q36Accep == 1 & Q37WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(weight)]

* LFS 2007a
gen sss = 0
replace sss = 1 if Q38LookW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q38BgnBu == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q32YnotW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 6 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 8 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 9 & Q36Accep == 1 & Q37WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(weight)]

* LFS 2006b
gen sss = 0
replace sss = 1 if Q38LookW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q38BgnBu == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q32YnotW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 6 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 8 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 9 & Q36Accep == 1 & Q37WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(weight)]

* LFS 2006a
gen sss = 0
replace sss = 1 if Q38LookW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q38BgnBu == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q32YnotW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 6 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 8 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 9 & Q36Accep == 1 & Q37WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(weight)]

* LFS 2005b
gen sss = 0
replace sss = 1 if Q38LookW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q38BgnBu == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 1 if Q32YnotW == 1 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 6 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 8 & Q36Accep == 1 & Q37WhnSt == 1
replace sss = 2 if Q311RsnN == 9 & Q36Accep == 1 & Q37WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(pweight)]
```

Table B.2: Continued

```
* LFS 2005a
gen sss = 0
replace sss = 1 if Q37LookW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q37BgnBu == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q31YnotW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 6 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 8 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 9 & Q35Accep == 1 & Q36WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(pweight)]

* LFS 2004b
gen sss = 0
replace sss = 1 if Q37LookW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q37BgnBu == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q31YnotW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 6 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 8 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 9 & Q35Accep == 1 & Q36WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(pweight)]

* LFS 2004a
gen sss = 0
replace sss = 1 if Q37LookW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q37BgnBu == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q31YnotW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 6 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 8 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 9 & Q35Accep == 1 & Q36WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(pweight)]

* LFS 2003b
gen sss = 0
replace sss = 1 if Q37LookW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q37BgnBs == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q31YnotW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 6 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 8 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 9 & Q35Accep == 1 & Q36WhnSt == 1
tab sss if Age >= 15 & Age <= 65 [w=int(pweight)]

* LFS 2003a
gen sss = 0
replace sss = 1 if Q37LookW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q37BgnBu == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q31YnotW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 6 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 8 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 9 & Q35Accep == 1 & Q36WhnSt == 1
tab sss if Age >= 15 & Age <= 65[w=int(pweight)]

* LFS 2002b
gen sss = 0
replace sss = 1 if Q37LookW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q37BgnBs == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 1 if Q31YnotW == 1 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 6 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 8 & Q35Accep == 1 & Q36WhnSt == 1
replace sss = 2 if Q310RsnN == 9 & Q35Accep == 1 & Q36WhnSt == 1
tab sss if D_Age >= 15 & D_Age <= 65 [w=int(pweight)]
```

Table B.2: Continued

```
* LFS 2002a
gen sss = 0
replace sss = 1 if Q34aLook == 1 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 1 if Q34bBgnB == 1 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 1 if Q31YnotW == 1 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 2 if Q37YnotS == 6 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 2 if Q37YnotS == 8 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 2 if Q37YnotS == 9 & Q32Accep == 1 & Q33Whnst == 1
tab sss if D_Age >= 15 & D_Age <= 65 [w=int(pweight)]

* LFS 2001b
gen sss = 0
replace sss = 1 if Q34aLook == 1 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 1 if Q34bBgnB == 1 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 1 if Q31YnotW == 1 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 2 if Q37YnotS == 6 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 2 if Q37YnotS == 8 & Q32Accep == 1 & Q33Whnst == 1
replace sss = 2 if Q37YnotS == 9 & Q32Accep == 1 & Q33Whnst == 1
tab sss if D_Age >= 15 & D_Age <= 65 [w=int(pweight)]

* LFS 2001a
gen sss = 0
replace sss = 1 if Q34aLook == 1 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 1 if Q34bBgnB == 1 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 1 if Q31YnotW == 1 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 2 if Q37YnotL == 6 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 2 if Q37YnotL == 8 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 2 if Q37YnotL == 9 & Q32Accep == 1 & Q33WhenS == 1
tab sss if D_Age >= 15 & D_Age <= 65 [w=int(pweight)]

* LFS 2000b
gen sss = 0
replace sss = 1 if Q34aLook == 1 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 1 if Q34bBgnB == 1 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 1 if Q31YnotW == 1 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 2 if Q37YnotL == 6 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 2 if Q37YnotL == 8 & Q32Accep == 1 & Q33WhenS == 1
replace sss = 2 if Q37YnotL == 9 & Q32Accep == 1 & Q33WhenS == 1
tab sss if D_Age >= 15 & D_Age <= 65 [w=int(pweight)]

* LFS 2000a
gen sss = 0
replace sss = 1 if q37_lookw == 1 & q35_accep == 1 & q36_whnst == 1
replace sss = 1 if q37_bgnbu == 1 & q35_accep == 1 & q36_whnst == 1
replace sss = 1 if q31_ynotw == 1 & q35_accep == 1 & q36_whnst == 1
replace sss = 2 if q310_rsnn == 6 & q35_accep == 1 & q36_whnst == 1
replace sss = 2 if q310_rsnn == 8 & q35_accep == 1 & q36_whnst == 1
replace sss = 2 if q310_rsnn == 9 & q35_accep == 1 & q36_whnst == 1
tab sss if age >= 15 & age <= 65 [w=int(weight)]
```

Note: sss: 0 – inactive or employed; 1 – unemployed; 2 – discouraged workseekers

Appendix C: Poverty and inequality estimates in each survey

Table C.1: Poverty headcount ratios at the three poverty lines in each survey

Survey	Per capita variable	Year	Poverty headcount ratio R211	Poverty headcount ratio R322	Poverty headcount ratio R593
Census/ CS	Income – No imputations	1996	0.501	0.606	0.728
		2001	0.568	0.670	0.789
		2007	0.397	0.529	0.694
	Income – After SRMI1	1996	0.493	0.601	0.726
		2001	0.547	0.647	0.769
		2007	0.351	0.478	0.656
	Income – After SRMI2	1996	0.441	0.576	0.715
		2001	0.447	0.592	0.750
		2007	0.329	0.463	0.650
IES	Income – STC	1995	0.286	0.434	0.622
		2000	0.429	0.559	0.710
		2005/2006	0.338	0.488	0.657
	Expenditure – STC	1995	0.300	0.447	0.629
		2000	0.430	0.564	0.714
		2005/2006	0.303	0.466	0.654
	Income - COICOP	1995	0.318	0.462	0.642
		2000	0.442	0.572	0.723
		2005/2006	0.316	0.473	0.652
	Consumption - COICOP	1995	0.339	0.502	0.691
		2000	0.458	0.601	0.753
		2005/2006	0.320	0.500	0.699
OHS	Expenditure – No imputations	1996	0.588	0.704	0.815
		1997	0.665	0.768	0.875
		1998	0.667	0.781	0.871
		1999	0.652	0.742	0.838
	Income – No imputations	1999	0.518	0.617	0.745
	Expenditure – After SRMI2	1996	0.565	0.687	0.816
		1997	0.660	0.764	0.870
		1998	0.656	0.771	0.865
		1999	0.634	0.727	0.829
	Income – After SRMI2	1999	0.494	0.596	0.729
LFS	Expenditure – No imputations	2001	0.693	0.773	0.859
		2002	0.684	0.788	0.853
		2003	0.678	0.758	0.838
		2004	0.649	0.738	0.827
	Expenditure – After SRMI2	2001	0.682	0.764	0.852
		2002	0.674	0.779	0.845
		2003	0.669	0.750	0.830
		2004	0.639	0.730	0.820

Table C.1: Continued

Survey	Per capita variable	Year	Poverty headcount ratio		
			R211	**R322**	**R593**
GHS	Expenditure – No imputations	2002	0.689	0.778	0.861
		2003	0.681	0.762	0.845
		2004	0.637	0.733	0.823
		2005	0.618	0.710	0.840
		2006	0.619	0.731	0.842
		2007	0.614	0.695	0.822
		2008	0.618	0.712	0.829
		2009	0.552	0.675	0.790
	Expenditure – After SRMI2	2002	0.677	0.768	0.854
		2003	0.668	0.751	0.837
		2004	0.627	0.723	0.815
		2005	0.612	0.705	0.836
		2006	0.615	0.728	0.839
		2007	0.611	0.692	0.820
		2008	0.610	0.706	0.824
		2009	0.549	0.674	0.790
PSLSD	Income	1993	0.475	0.598	0.745
	Expenditure	1993	0.398	0.566	0.750
NIDS	Income	2008	0.302	0.471	0.656
	Expenditure	2008	0.386	0.532	0.687
AMPS	Income	1993	0.438	0.586	0.737
		1994	0.439	0.593	0.735
		1995	0.464	0.594	0.741
		1996	0.473	0.610	0.744
		1997	0.456	0.589	0.732
		1998	0.453	0.583	0.725
		1999	0.469	0.591	0.723
		2000	0.458	0.582	0.723
		2001	0.466	0.579	0.717
		2002	0.434	0.563	0.709
		2003	0.418	0.554	0.704
		2004	0.415	0.548	0.703
		2005	0.391	0.519	0.680
		2006	0.385	0.512	0.673
		2007	0.332	0.455	0.613
		2008	0.292	0.410	0.580
		2009	0.306	0.414	0.574

Table C.2: Poverty headcount ratios by race (Poverty line: R322 per month, 2000 prices)

Survey	Variable	Year	Black	Coloured	Indian	White	All
Census/ CS	Income – No imputations	1996	0.720	0.393	0.158	0.055	0.606
		2001	0.765	0.453	0.189	0.068	0.670
		2007	0.609	0.320	0.145	0.050	0.529
	Total income – After SRMI1	1996	0.718	0.385	0.151	0.062	0.601
		2001	0.755	0.436	0.177	0.072	0.647
		2007	0.562	0.282	0.126	0.059	0.478
	Total income – After SRMI2	1996	0.693	0.365	0.125	0.033	0.576
		2001	0.701	0.369	0.119	0.024	0.592
		2007	0.551	0.256	0.098	0.016	0.463
IES	Total income – Standard Trade Classification	1995	0.533	0.289	0.034	0.007	0.434
		2000	0.660	0.347	0.128	0.072	0.559
		2005/2006	0.576	0.289	0.141	0.019	0.488
	Total expenditure – Standard Trade Classification	1995	0.548	0.313	0.028	0.006	0.447
		2000	0.673	0.368	0.101	0.013	0.564
		2005/2006	0.545	0.346	0.094	0.005	0.466
	Total income - COICOP	1995	0.566	0.322	0.034	0.009	0.462
		2000	0.674	0.370	0.130	0.079	0.572
		2005/2006	0.559	0.275	0.136	0.015	0.473
	Total consumption - COICOP	1995	0.611	0.382	0.067	0.010	0.502
		2000	0.713	0.412	0.120	0.025	0.601
		2005/2006	0.587	0.358	0.090	0.004	0.500
OHS	Total expenditure – No imputations	1996	0.823	0.596	0.311	0.102	0.704
		1997	0.885	0.672	0.301	0.085	0.768
		1998	0.895	0.677	0.384	0.110	0.781
		1999	0.845	0.621	0.304	0.105	0.742
	Total income – No imputations	1999	0.720	0.408	0.166	0.033	0.617
	Total expenditure – After SRMI2	1996	0.741	0.556	0.524	0.406	0.687
		1997	0.886	0.670	0.290	0.079	0.764
		1998	0.894	0.666	0.367	0.099	0.771
		1999	0.843	0.603	0.274	0.088	0.727
	Total income – After SRMI2	1999	0.712	0.382	0.146	0.026	0.596
LFS	Total expenditure – No imputations	2001	0.879	0.695	0.253	0.096	0.773
		2002	0.888	0.720	0.306	0.096	0.788
		2003	0.865	0.649	0.217	0.055	0.758
		2004	0.843	0.609	0.210	0.061	0.738
	Total expenditure – After SRMI2	2001	0.877	0.685	0.240	0.088	0.764
		2002	0.887	0.714	0.300	0.086	0.779
		2003	0.864	0.648	0.214	0.050	0.750
		2004	0.842	0.600	0.222	0.056	0.730
GHS	Total expenditure – No Imputations	2002	0.875	0.689	0.324	0.097	0.778
		2003	0.857	0.655	0.319	0.079	0.762
		2004	0.837	0.595	0.225	0.060	0.733
		2005	0.817	0.542	0.224	0.055	0.710
		2006	0.833	0.567	0.371	0.081	0.731
		2007	0.796	0.541	0.284	0.060	0.695
		2008	0.816	0.542	0.188	0.088	0.712
		2009	0.775	0.528	0.232	0.059	0.675

Table C.2: Continued

Survey	Variable	Year	Black	Coloured	Indian	White	All
GHS	Total expenditure – After SRMI2	2002	0.874	0.682	0.305	0.088	0.768
		2003	0.855	0.645	0.316	0.070	0.751
		2004	0.835	0.587	0.212	0.054	0.723
		2005	0.816	0.537	0.222	0.052	0.705
		2006	0.833	0.564	0.368	0.078	0.728
		2007	0.795	0.538	0.281	0.057	0.692
		2008	0.815	0.536	0.192	0.081	0.706
		2009	0.776	0.522	0.228	0.055	0.674
PSLSD	Total income	1993	0.735	0.298	0.094	0.025	0.598
	Total expenditure	1993	0.699	0.343	0.026	0.005	0.566
NIDS	Total income	2008	0.560	0.282	0.108	0.014	0.471
	Total expenditure	2008	0.626	0.339	0.170	0.029	0.532
AMPS	Total income	1993	0.728	0.391	0.107	0.008	0.586
		1994	0.736	0.385	0.101	0.012	0.593
		1995	0.736	0.351	0.109	0.008	0.594
		1996	0.747	0.385	0.126	0.011	0.610
		1997	0.725	0.354	0.111	0.012	0.589
		1998	0.726	0.346	0.116	0.010	0.583
		1999	0.731	0.367	0.122	0.011	0.591
		2000	0.707	0.364	0.098	0.014	0.582
		2001	0.707	0.347	0.104	0.015	0.579
		2002	0.678	0.381	0.123	0.019	0.563
		2003	0.672	0.350	0.104	0.020	0.554
		2004	0.661	0.346	0.098	0.015	0.548
		2005	0.622	0.368	0.079	0.024	0.519
		2006	0.616	0.343	0.071	0.020	0.512
		2007	0.550	0.303	0.057	0.016	0.455
		2008	0.500	0.234	0.062	0.008	0.410
		2009	0.503	0.268	0.043	0.014	0.414

Table C.3: Poverty gap indices at different poverty lines in each survey

Survey	Per capita variable	Year	Poverty gap ratio		
			R211	R322	R593
Census/ CS	Income – No imputations	1996	0.332	0.411	0.532
		2001	0.396	0.476	0.592
		2007	0.194	0.288	0.444
	Income – After SRMI1	1996	0.327	0.406	0.528
		2001	0.380	0.457	0.573
		2007	0.171	0.256	0.406
	Income – After SRMI2	1996	0.233	0.332	0.479
		2001	0.208	0.320	0.481
		2007	0.129	0.222	0.383
IES	Income – STC	1995	0.106	0.195	0.352
		2000	0.206	0.307	0.462
		2005/2006	0.137	0.234	0.395
	Expenditure – STC	1995	0.115	0.206	0.363
		2000	0.197	0.303	0.462
		2005/2006	0.110	0.207	0.376
	Income - COICOP	1995	0.130	0.222	0.377
		2000	0.216	0.318	0.474
		2005/2006	0.121	0.217	0.382
	Consumption - COICOP	1995	0.130	0.232	0.405
		2000	0.212	0.324	0.491
		2005/2006	0.112	0.216	0.400
OHS	Expenditure – No imputations	1996	0.308	0.425	0.583
		1997	0.361	0.486	0.640
		1998	0.390	0.507	0.654
		1999	0.335	0.457	0.610
	Income – No imputations	1999	0.250	0.357	0.507
	Expenditure – After SRMI2	1996	0.299	0.413	0.573
		1997	0.357	0.481	0.636
		1998	0.382	0.499	0.646
		1999	0.321	0.442	0.597
	Income – After SRMI2	1999	0.234	0.339	0.489
LFS	Expenditure – No imputations	2001	0.411	0.521	0.654
		2002	0.424	0.535	0.666
		2003	0.389	0.503	0.639
		2004	0.239	0.477	0.619
	Expenditure – After SRMI2	2001	0.402	0.512	0.646
		2002	0.416	0.526	0.657
		2003	0.383	0.496	0.632
		2004	0.361	0.469	0.611

Table C.3: Continued

Survey	Variable	Year	Poverty gap ratio		
			R211	**R322**	**R593**
GHS	Expenditure – No imputations	2002	0.421	0.532	0.667
		2003	0.392	0.505	0.643
		2004	0.347	0.464	0.610
		2005	0.342	0.455	0.601
		2006	0.347	0.465	0.613
		2007	0.316	0.436	0.591
		2008	0.329	0.446	0.598
		2009	0.285	0.398	0.553
	Expenditure – After SRMI2	2002	0.412	0.523	0.658
		2003	0.382	0.495	0.633
		2004	0.340	0.456	0.602
		2005	0.337	0.451	0.596
		2006	0.344	0.462	0.610
		2007	0.314	0.433	0.588
		2008	0.324	0.440	0.593
		2009	0.281	0.395	0.551
PSLSD	Income	1993	0.253	0.352	0.504
	Expenditure	1993	0.155	0.271	0.455
NIDS	Income	2008	0.122	0.215	0.380
	Expenditure	2008	0.153	0.261	0.427
AMPS	Income	1993	0.200	0.310	0.477
		1994	0.196	0.310	0.477
		1995	0.227	0.333	0.492
		1996	0.218	0.332	0.495
		1997	0.213	0.322	0.482
		1998	0.211	0.319	0.477
		1999	0.215	0.326	0.481
		2000	0.211	0.320	0.477
		2001	0.223	0.329	0.479
		2002	0.203	0.308	0.462
		2003	0.185	0.291	0.451
		2004	0.178	0.284	0.445
		2005	0.175	0.274	0.427
		2006	0.171	0.269	0.422
		2007	0.142	0.230	0.372
		2008	0.124	0.204	0.340
		2009	0.132	0.211	0.344

Table C.4: Poverty gap indices by race (Poverty line: R322 per month, 2000 prices)

Survey	Variable	Year	Black	Coloured	Indian	White	All
Census/ CS	Income – No imputations	1996	0.505	0.194	0.077	0.036	0.411
		2001	0.551	0.250	0.106	0.048	0.476
		2007	0.332	0.154	0.077	0.037	0.288
	Total income – After SRMI1	1996	0.493	0.191	0.074	0.044	0.406
		2001	0.542	0.238	0.101	0.053	0.457
		2007	0.300	0.135	0.068	0.048	0.256
	Total income – After SRMI2	1996	0.407	0.156	0.045	0.015	0.332
		2001	0.385	0.156	0.044	0.009	0.320
		2007	0.267	0.105	0.039	0.006	0.222
IES	Total income – Standard Trade Classification	1995	0.244	0.099	0.008	0.002	0.195
		2000	0.365	0.149	0.057	0.059	0.307
		2005/2006	0.279	0.114	0.060	0.011	0.234
	Total expenditure – Standard Trade Classification	1995	0.256	0.111	0.006	0.002	0.206
		2000	0.368	0.149	0.034	0.006	0.303
		2005/2006	0.245	0.134	0.029	0.001	0.207
	Total income - COICOP	1995	0.276	0.119	0.009	0.004	0.222
		2000	0.377	0.161	0.061	0.065	0.318
		2005/2006	0.260	0.103	0.053	0.008	0.217
	Total consumption - COICOP	1995	0.288	0.133	0.014	0.003	0.232
		2000	0.392	0.167	0.039	0.007	0.324
		2005/2006	0.256	0.139	0.024	0.001	0.216
OHS	Total expenditure – No imputations	1996	0.514	0.309	0.103	0.037	0.425
		1997	0.575	0.342	0.086	0.031	0.486
		1998	0.600	0.343	0.144	0.042	0.507
		1999	0.534	0.309	0.118	0.035	0.457
	Total income – No imputations	1999	0.425	0.175	0.064	0.012	0.357
	Total expenditure – After SRMI2	1996	0.456	0.294	0.281	0.215	0.413
		1997	0.573	0.338	0.082	0.028	0.481
		1998	0.596	0.330	0.137	0.037	0.499
		1999	0.526	0.292	0.103	0.029	0.442
	Total income – After SRMI2	1999	0.413	0.160	0.053	0.010	0.339
LFS	Total expenditure – No imputations	2001	0.606	0.401	0.101	0.031	0.521
		2002	0.618	0.406	0.124	0.032	0.535
		2003	0.587	0.343	0.082	0.018	0.503
		2004	0.558	0.306	0.073	0.022	0.477
	Total expenditure – After SRMI2	2001	0.600	0.390	0.096	0.028	0.512
		2002	0.615	0.394	0.121	0.028	0.526
		2003	0.585	0.341	0.081	0.016	0.496
		2004	0.556	0.296	0.078	0.020	0.469
GHS	Total expenditure – No Imputations	2002	0.614	0.399	0.124	0.032	0.532
		2003	0.584	0.348	0.107	0.023	0.505
		2004	0.544	0.294	0.081	0.015	0.464
		2005	0.535	0.280	0.088	0.018	0.455
		2006	0.543	0.300	0.128	0.027	0.465
		2007	0.510	0.277	0.113	0.024	0.436
		2008	0.521	0.292	0.075	0.027	0.446
		2009	0.466	0.255	0.078	0.021	0.398

Table C.4: Continued

Survey	Variable	Year	Black	Coloured	Indian	White	All
GHS	Total expenditure – After SRMI2	2002	0.610	0.387	0.117	0.029	0.523
		2003	0.579	0.338	0.102	0.020	0.495
		2004	0.540	0.288	0.076	0.013	0.456
		2005	0.533	0.277	0.086	0.017	0.451
		2006	0.541	0.298	0.126	0.025	0.462
		2007	0.509	0.274	0.112	0.023	0.433
		2008	0.518	0.287	0.075	0.025	0.440
		2009	0.484	0.250	0.076	0.020	0.395
PSLSD	Total income	1993	0.438	0.127	0.036	0.024	0.352
	Total expenditure	1993	0.340	0.120	0.007	0.001	0.271
NIDS	Total income	2008	0.259	0.104	0.032	0.009	0.215
	Total expenditure	2008	0.315	0.122	0.026	0.008	0.261
AMPS	Total income	1993	0.392	0.158	0.033	0.003	0.310
		1994	0.392	0.140	0.031	0.004	0.310
		1995	0.419	0.143	0.032	0.003	0.333
		1996	0.414	0.151	0.036	0.003	0.332
		1997	0.403	0.142	0.034	0.004	0.322
		1998	0.405	0.133	0.040	0.003	0.319
		1999	0.411	0.148	0.034	0.003	0.326
		2000	0.395	0.149	0.033	0.005	0.320
		2001	0.406	0.159	0.038	0.006	0.329
		2002	0.376	0.167	0.044	0.007	0.308
		2003	0.358	0.148	0.036	0.007	0.291
		2004	0.347	0.144	0.035	0.006	0.284
		2005	0.334	0.155	0.027	0.009	0.274
		2006	0.328	0.148	0.025	0.008	0.269
		2007	0.282	0.127	0.021	0.007	0.230
		2008	0.251	0.099	0.023	0.002	0.204
		2009	0.259	0.120	0.017	0.005	0.211

Table C.5: Squared poverty gap indices at different poverty lines in each survey

Survey	Per capita variable	Year	Squared poverty gap ratio		
			R211	R322	R593
Census/ CS	Income – No imputations	1996	0.263	0.328	0.437
		2001	0.330	0.393	0.501
		2007	0.133	0.197	0.327
	Income – After SRMI1	1996	0.259	0.323	0.432
		2001	0.316	0.377	0.483
		2007	0.118	0.175	0.294
	Income – After SRMI2	1996	0.148	0.228	0.363
		2001	0.120	0.206	0.355
		2007	0.071	0.133	0.263
IES	Income – STC	1995	0.053	0.111	0.236
		2000	0.127	0.204	0.342
		2005/2006	0.075	0.141	0.275
	Expenditure – STC	1995	0.058	0.119	0.246
		2000	0.115	0.196	0.339
		2005/2006	0.054	0.116	0.251
	Income - COICOP	1995	0.069	0.133	0.260
		2000	0.135	0.214	0.353
		2005/2006	0.063	0.126	0.260
	Consumption - COICOP	1995	0.066	0.134	0.276
		2000	0.125	0.210	0.362
		2005/2006	0.054	0.120	0.265
OHS	Expenditure – No imputations	1996	0.197	0.299	0.456
		1997	0.235	0.348	0.512
		1998	0.261	0.374	0.532
		1999	0.222	0.327	0.486
	Income – No imputations	1999	0.160	0.246	0.389
	Expenditure – After SRMI2	1996	0.195	0.291	0.445
		1997	0.231	0.343	0.508
		1998	0.255	0.366	0.523
		1999	0.211	0.314	0.472
	Income – After SRMI2	1999	0.148	0.232	0.371
LFS	Expenditure – No imputations	2001	0.289	0.395	0.542
		2002	0.300	0.407	0.554
		2003	0.265	0.373	0.524
		2004	0.241	0.347	0.500
	Expenditure – After SRMI2	2001	0.282	0.386	0.533
		2002	0.293	0.400	0.546
		2003	0.261	0.368	0.517
		2004	0.236	0.341	0.493

Table C.5: Continued

Survey	Variable	Year	Squared poverty gap ratio		
			R211	R322	R593
GHS	Expenditure – No imputations	2002	0.295	0.404	0.553
		2003	0.268	0.376	0.527
		2004	0.227	0.334	0.489
		2005	0.226	0.329	0.481
		2006	0.227	0.334	0.490
		2007	0.201	0.306	0.464
		2008	0.211	0.317	0.474
		2009	0.179	0.276	0.429
	Expenditure – After SRMI2	2002	0.288	0.395	0.544
		2003	0.261	0.367	0.517
		2004	0.222	0.327	0.481
		2005	0.222	0.325	0.476
		2006	0.225	0.332	0.487
		2007	0.199	0.304	0.462
		2008	0.207	0.312	0.468
		2009	0.176	0.273	0.426
PSLSD	Income	1993	0.184	0.255	0.388
	Expenditure	1993	0.080	0.159	0.316
NIDS	Income	2008	0.067	0.127	0.258
	Expenditure	2008	0.079	0.156	0.301
AMPS	Income	1993	0.118	0.200	0.349
		1994	0.114	0.197	0.348
		1995	0.140	0.223	0.368
		1996	0.129	0.216	0.368
		1997	0.126	0.210	0.358
		1998	0.124	0.209	0.354
		1999	0.125	0.212	0.359
		2000	0.123	0.208	0.354
		2001	0.134	0.219	0.360
		2002	0.120	0.201	0.342
		2003	0.106	0.185	0.328
		2004	0.099	0.178	0.321
		2005	0.101	0.175	0.309
		2006	0.098	0.171	0.305
		2007	0.078	0.142	0.264
		2008	0.069	0.125	0.237
		2009	0.074	0.132	0.243

Table C.6: Squared poverty gap indices by race (Poverty line: R322 per month, 2000 prices)

Survey	Variable	Year	Black	Coloured	Indian	White	All
Census/ CS	Income – No imputations	1996	0.405	0.131	0.055	0.030	0.328
		2001	0.458	0.181	0.081	0.042	0.393
		2007	0.228	0.101	0.055	0.033	0.197
	Total income – After SRMI1	1996	0.395	0.129	0.054	0.039	0.323
		2001	0.450	0.172	0.078	0.047	0.377
		2007	0.204	0.089	0.049	0.045	0.175
	Total income – After SRMI2	1996	0.282	0.090	0.025	0.010	0.228
		2001	0.250	0.086	0.022	0.005	0.206
		2007	0.161	0.058	0.021	0.003	0.133
IES	Total income – Standard Trade Classification	1995	0.140	0.047	0.003	0.001	0.111
		2000	0.243	0.086	0.036	0.052	0.204
		2005/2006	0.169	0.063	0.037	0.007	0.141
	Total expenditure – Standard Trade Classification	1995	0.150	0.053	0.002	0.001	0.119
		2000	0.240	0.080	0.015	0.004	0.196
		2005/2006	0.138	0.070	0.015	0.001	0.116
	Total income - COICOP	1995	0.167	0.059	0.003	0.002	0.133
		2000	0.253	0.094	0.040	0.058	0.214
		2005/2006	0.151	0.054	0.030	0.005	0.126
	Total consumption - COICOP	1995	0.169	0.063	0.005	0.001	0.134
		2000	0.257	0.090	0.017	0.004	0.210
		2005/2006	0.142	0.071	0.013	0.000	0.120
OHS	Total expenditure – No imputations	1996	0.365	0.198	0.049	0.021	0.299
		1997	0.417	0.211	0.036	0.017	0.348
		1998	0.447	0.218	0.080	0.024	0.374
		1999	0.386	0.194	0.065	0.017	0.327
	Total income – No imputations	1999	0.296	0.103	0.037	0.008	0.246
	Total expenditure – After SRMI2	1996	0.325	0.196	0.181	0.141	0.291
		1997	0.414	0.208	0.034	0.015	0.343
		1998	0.443	0.208	0.075	0.022	0.366
		1999	0.378	0.181	0.055	0.014	0.314
	Total income – After SRMI2	1999	0.284	0.093	0.030	0.006	0.232
LFS	Total expenditure – No imputations	2001	0.463	0.277	0.053	0.016	0.395
		2002	0.475	0.278	0.070	0.016	0.407
		2003	0.441	0.220	0.044	0.008	0.373
		2004	0.411	0.190	0.038	0.011	0.347
	Total expenditure – After SRMI2	2001	0.458	0.268	0.051	0.014	0.386
		2002	0.472	0.267	0.068	0.014	0.400
		2003	0.438	0.218	0.043	0.007	0.368
		2004	0.408	0.183	0.040	0.010	0.341
GHS	Total expenditure – No Imputations	2002	0.470	0.273	0.064	0.015	0.404
		2003	0.439	0.228	0.051	0.010	0.376
		2004	0.395	0.180	0.041	0.006	0.334
		2005	0.390	0.177	0.043	0.008	0.329
		2006	0.394	0.195	0.065	0.013	0.334
		2007	0.362	0.170	0.059	0.012	0.306
		2008	0.373	0.188	0.040	0.013	0.317
		2009	0.327	0.159	0.039	0.010	0.276

Table C.6: Continued

Survey	Variable	Year	Black	Coloured	Indian	White	All
GHS	Total expenditure – After SRMI2	2002	0.466	0.262	0.060	0.014	0.395
		2003	0.434	0.219	0.048	0.009	0.367
		2004	0.392	0.175	0.038	0.005	0.327
		2005	0.388	0.175	0.042	0.008	0.325
		2006	0.393	0.193	0.064	0.012	0.332
		2007	0.361	0.168	0.058	0.012	0.304
		2008	0.371	0.184	0.040	0.012	0.312
		2009	0.324	0.153	0.038	0.010	0.273
PSLSD	Total income	1993	0.310	0.075	0.022	0.092	0.255
	Total expenditure	1993	0.202	0.058	0.003	0.000	0.159
NIDS	Total income	2008	0.154	0.052	0.016	0.006	0.127
	Total expenditure	2008	0.190	0.060	0.005	0.003	0.156
AMPS	Total income	1993	0.254	0.087	0.015	0.002	0.200
		1994	0.251	0.074	0.014	0.002	0.197
		1995	0.284	0.076	0.014	0.001	0.223
		1996	0.273	0.080	0.016	0.001	0.216
		1997	0.266	0.075	0.015	0.002	0.210
		1998	0.267	0.068	0.019	0.001	0.209
		1999	0.270	0.079	0.014	0.001	0.212
		2000	0.259	0.081	0.017	0.002	0.208
		2001	0.272	0.092	0.021	0.004	0.219
		2002	0.247	0.094	0.023	0.004	0.201
		2003	0.229	0.084	0.018	0.004	0.185
		2004	0.219	0.078	0.018	0.003	0.178
		2005	0.214	0.085	0.014	0.005	0.175
		2006	0.210	0.083	0.012	0.004	0.171
		2007	0.175	0.070	0.011	0.005	0.142
		2008	0.155	0.056	0.012	0.001	0.125
		2009	0.163	0.070	0.009	0.002	0.132

Table C.7: Gini coefficients by race in each survey

Survey	Variable	Year	Black	Coloured	Indian	White	All
Census/ CS	Income – No imputations	1996	0.698	0.557	0.510	0.480	0.742
		2001	0.781	0.659	0.628	0.620	0.825
		2007	0.700	0.653	0.657	0.603	0.774
	Total income – After SRMI1	1996	0.693	0.550	0.501	0.477	0.734
		2001	0.778	0.644	0.616	0.605	0.817
		2007	0.690	0.636	0.620	0.583	0.759
	Total income – After SRMI2	1996	0.620	0.528	0.481	0.459	0.694
		2001	0.654	0.601	0.582	0.566	0.756
		2007	0.663	0.615	0.608	0.559	0.743
IES	Total income – Standard Trade Classification	1995	0.564	0.488	0.472	0.438	0.655
		2000	0.630	0.555	0.519	0.510	0.711
		2005/2006	0.615	0.593	0.559	0.512	0.717
	Total expenditure – Standard Trade Classification	1995	0.569	0.499	0.463	0.434	0.660
		2000	0.610	0.548	0.486	0.484	0.710
		2005/2006	0.620	0.638	0.687	0.520	0.733
	Total income - COICOP	1995	0.571	0.493	0.469	0.442	0.660
		2000	0.628	0.553	0.513	0.511	0.709
		2005/2006	0.538	0.565	0.519	0.438	0.716
	Total consumption - COICOP	1995	0.527	0.452	0.453	0.436	0.612
		2000	0.550	0.477	0.434	0.433	0.651
		2005/2006	0.599	0.591	0.555	0.496	0.670
OHS	Total expenditure – No imputations	1996	0.590	0.512	0.452	0.465	0.646
		1997	0.525	0.507	0.409	0.510	0.663
		1998	0.561	0.488	0.421	0.476	0.662
		1999	0.596	0.558	0.564	0.562	0.713
	Total income – No imputations	1999	0.702	0.694	0.702	0.590	0.815
	Total expenditure – After SRMI2	1996	0.628	0.518	0.627	0.611	0.636
		1997	0.521	0.503	0.406	0.496	0.660
		1998	0.557	0.479	0.411	0.461	0.659
		1999	0.584	0.537	0.527	0.534	0.702
	Total income – After SRMI2	1999	0.692	0.688	0.683	0.571	0.815
LFS	Total expenditure – No imputations	2001	0.621	0.609	0.531	0.575	0.745
		2002	0.673	0.652	0.647	0.610	0.781
		2003	0.667	0.641	0.650	0.641	0.813
		2004	0.686	0.651	0.622	0.649	0.815
	Total expenditure – After SRMI2	2001	0.616	0.598	0.517	0.555	0.739
		2002	0.669	0.638	0.644	0.597	0.779
		2003	0.670	0.645	0.659	0.643	0.821
		2004	0.683	0.646	0.617	0.648	0.815
GHS	Total expenditure – No imputations	2002	0.621	0.605	0.564	0.556	0.736
		2003	0.679	0.628	0.649	0.598	0.772
		2004	0.613	0.559	0.602	0.524	0.720
		2005	0.631	0.562	0.597	0.578	0.737
		2006	0.638	0.592	0.633	0.609	0.753
		2007	0.641	0.640	0.627	0.567	0.735
		2008	0.683	0.724	0.664	0.604	0.787
		2009	0.735	0.749	0.690	0.583	0.815

Table C.7: Continued

Survey	Variable	Year	Black	Coloured	Indian	White	All
GHS	Total expenditure – After SRMI2	2002	0.619	0.594	0.555	0.550	0.736
		2003	0.674	0.615	0.641	0.591	0.771
		2004	0.611	0.556	0.589	0.521	0.723
		2005	0.630	0.558	0.594	0.575	0.738
		2006	0.634	0.586	0.626	0.601	0.748
		2007	0.640	0.637	0.624	0.564	0.735
		2008	0.682	0.720	0.666	0.597	0.787
		2009	0.719	0.733	0.676	0.576	0.806
PSLSD	Total income	1993	0.559	0.441	0.479	0.451	0.696
	Total expenditure	1993	0.449	0.405	0.371	0.322	0.595
NIDS	Total income	2008	0.594	0.531	0.549	0.484	0.683
	Total expenditure	2008	0.589	0.565	0.528	0.459	0.685
AMPS	Total income	1993	0.513	0.493	0.415	0.390	0.672
		1994	0.510	0.468	0.393	0.393	0.665
		1995	0.538	0.469	0.426	0.387	0.674
		1996	0.532	0.479	0.413	0.392	0.678
		1997	0.543	0.479	0.433	0.391	0.674
		1998	0.547	0.471	0.436	0.406	0.683
		1999	0.556	0.486	0.465	0.405	0.685
		2000	0.582	0.502	0.458	0.421	0.682
		2001	0.585	0.517	0.472	0.420	0.685
		2002	0.581	0.523	0.468	0.435	0.670
		2003	0.577	0.517	0.506	0.477	0.686
		2004	0.570	0.516	0.506	0.457	0.678
		2005	0.592	0.537	0.474	0.510	0.683
		2006	0.594	0.531	0.481	0.520	0.685
		2007	0.592	0.498	0.486	0.497	0.660
		2008	0.601	0.506	0.533	0.521	0.666
		2009	0.618	0.503	0.449	0.448	0.644

Figure C.1: Proportion of households reporting monthly household income or expenditure of below R1 200, AMPSs, OHSs, LFSs and GHSs

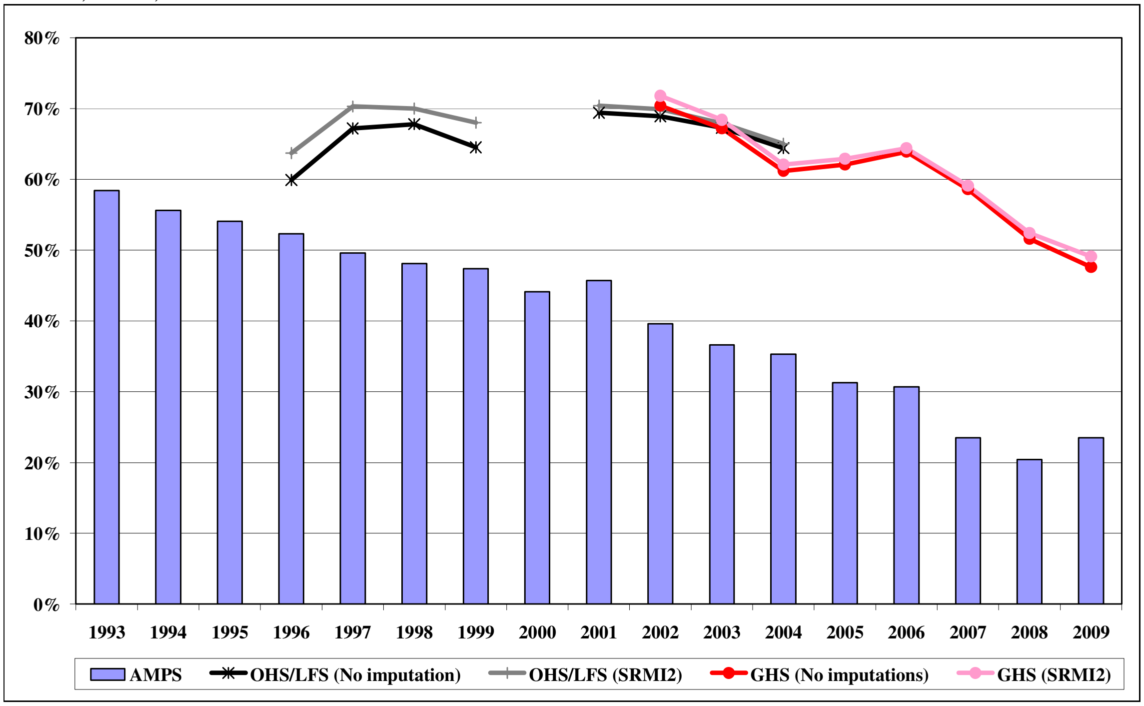


Figure C.2: Proportion of households reporting monthly household income or expenditure of below R800, AMPSs, OHSs, LFSs and GHSs

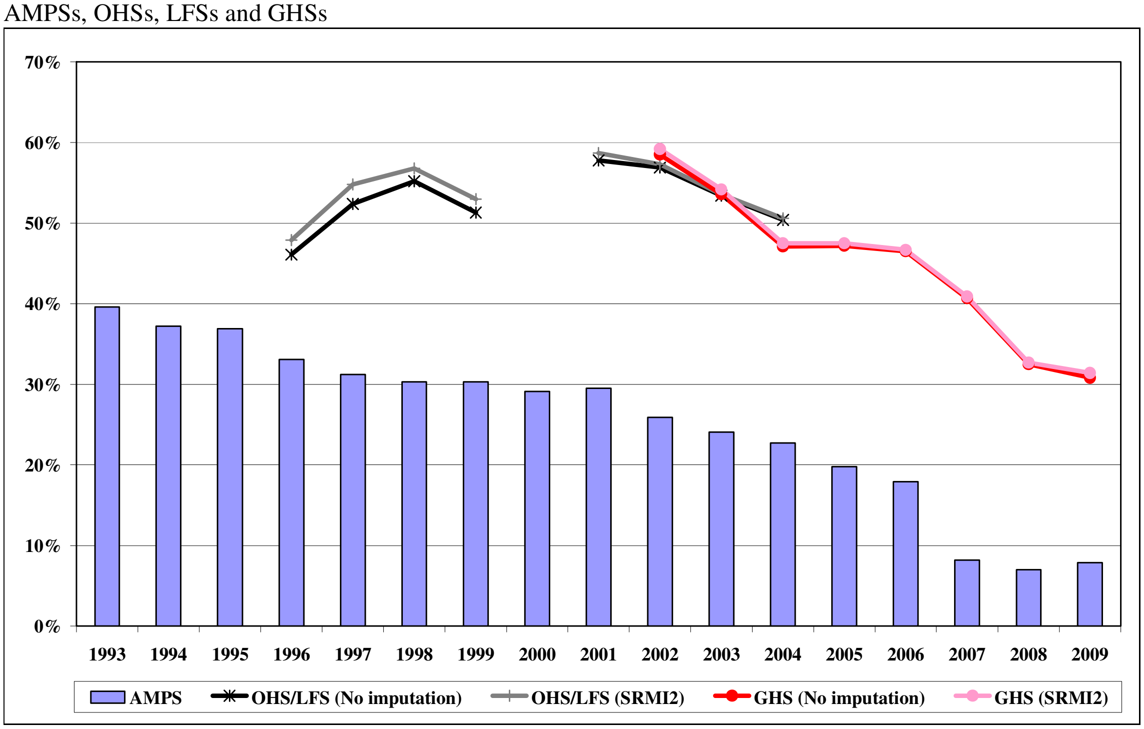


Figure C.3: Poverty headcount ratios at different poverty lines: Census (No imputations)

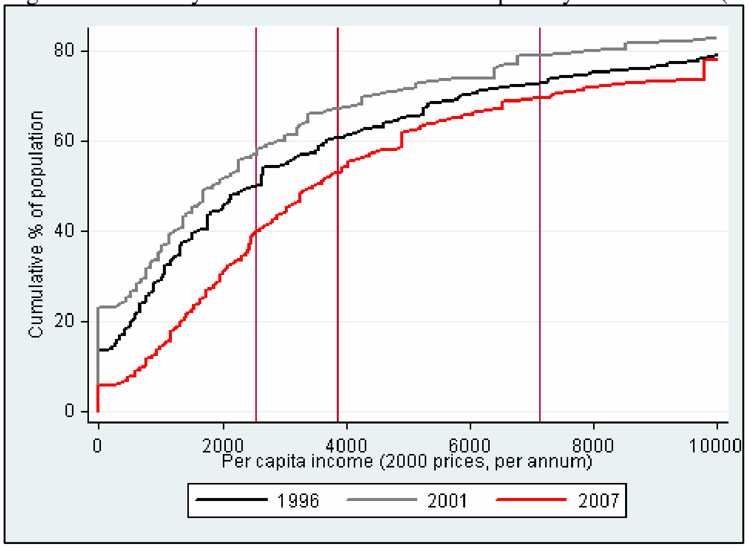


Figure C.4: Poverty headcount ratios at different poverty lines: Census (After SRMI1)

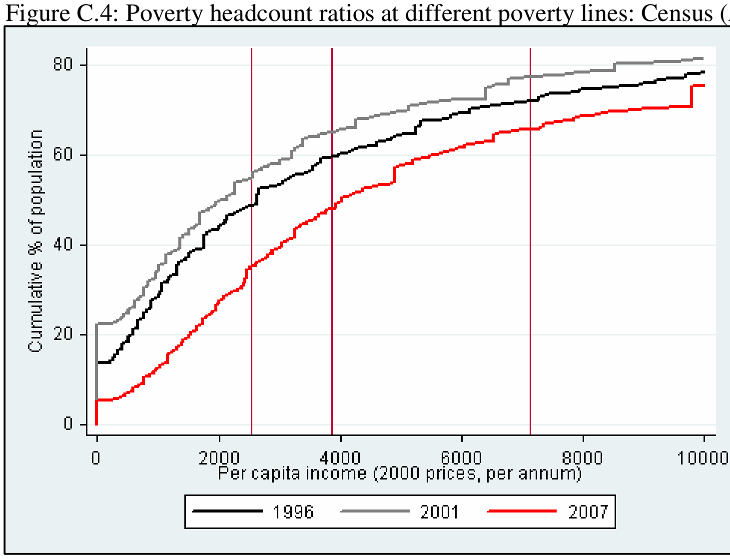


Figure C.5: Poverty headcount ratios at different poverty lines: Census (After SRMI2)

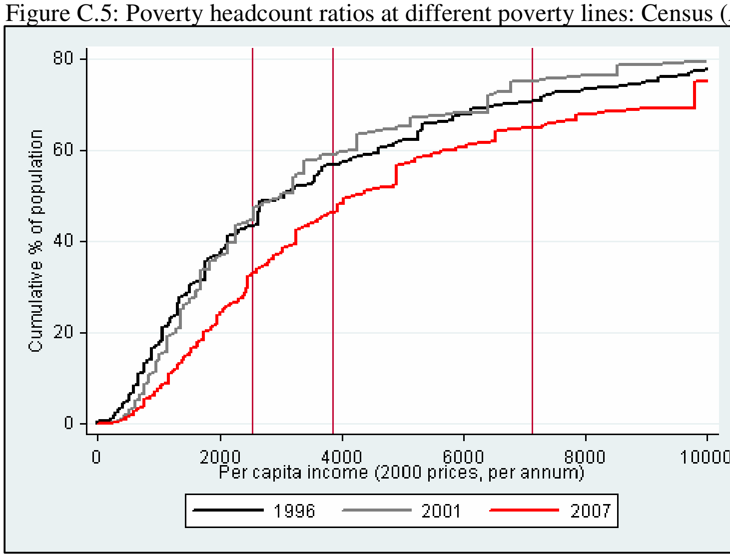


Figure C.6: Poverty headcount ratios at different poverty lines: IES Income (STC)

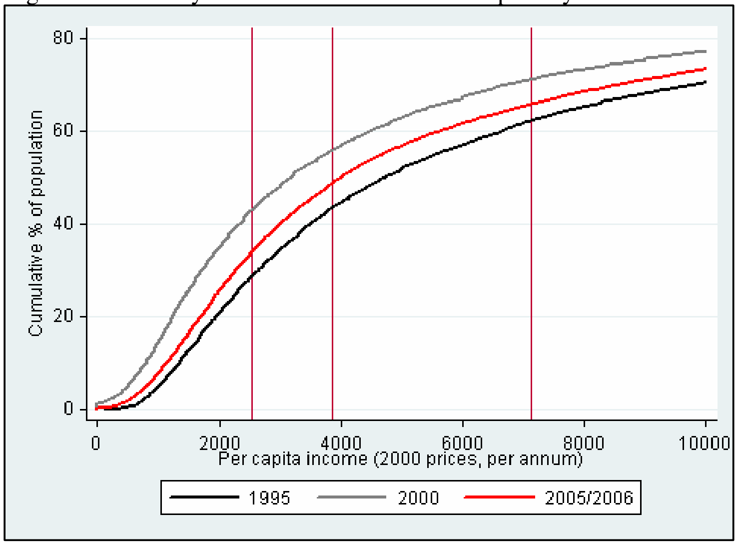


Figure C.7: Poverty headcount ratios at different poverty lines: IES Expenditure (STC)

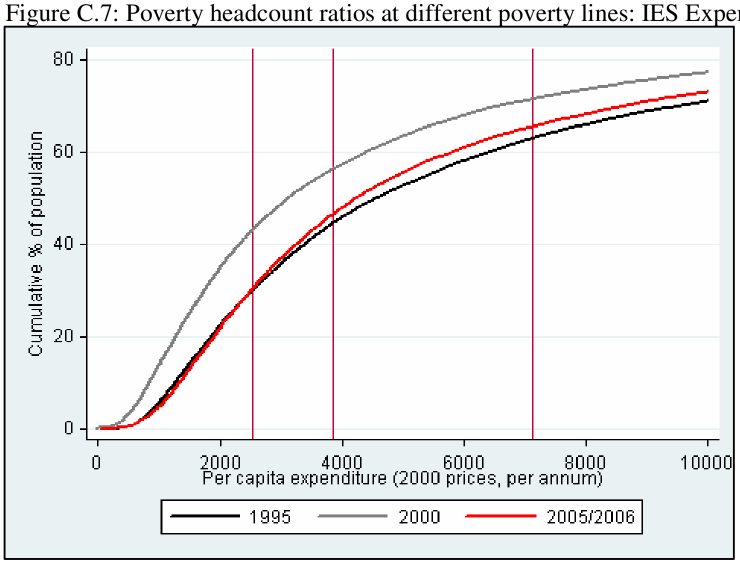


Figure C.8: Poverty headcount ratios at different poverty lines: IES Income (COICOP)

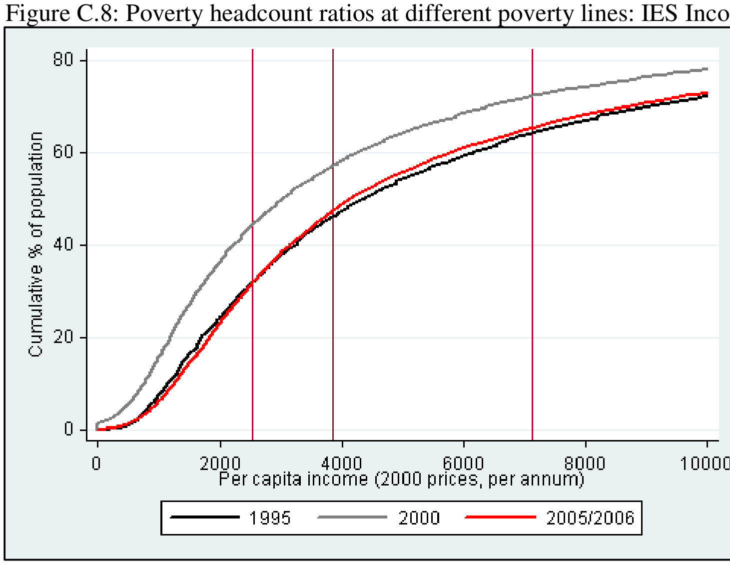


Figure C.9: Poverty headcount ratios at different poverty lines: IES Consumption (COICOP)

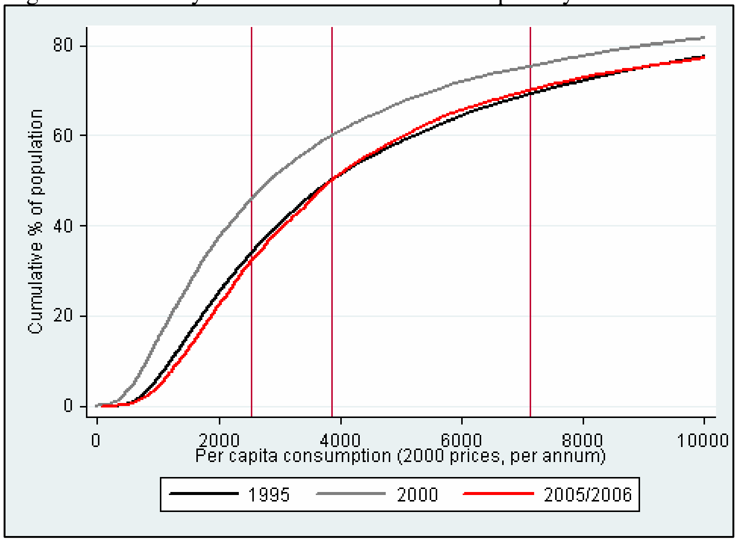


Figure C.10: Poverty headcount ratios at different poverty lines: Selected OHSs/LFSs (No imputations)

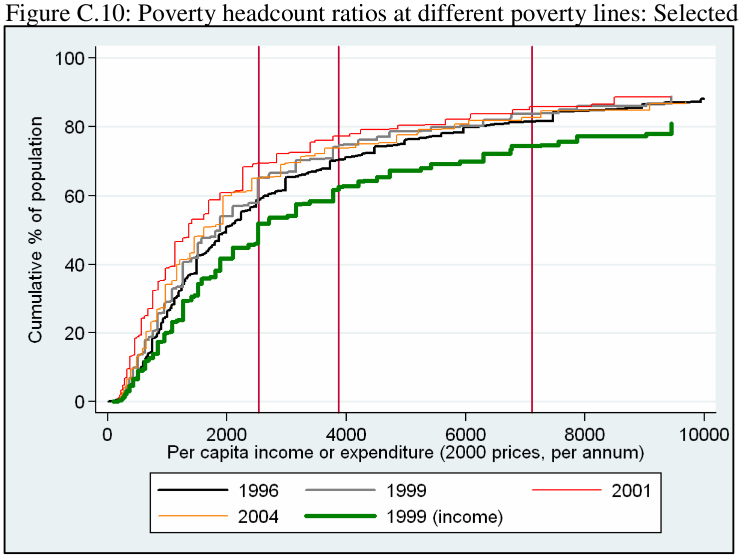


Figure C.11: Poverty headcount ratios at different poverty lines: Selected OHSs/LFSs (After SRMI2)

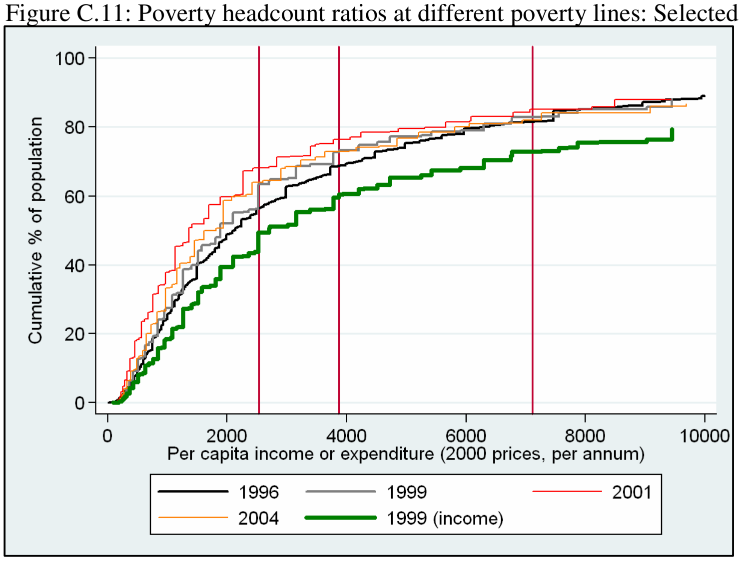


Figure C.12: Poverty headcount ratios at different poverty lines: Selected GHSs (No imputations)

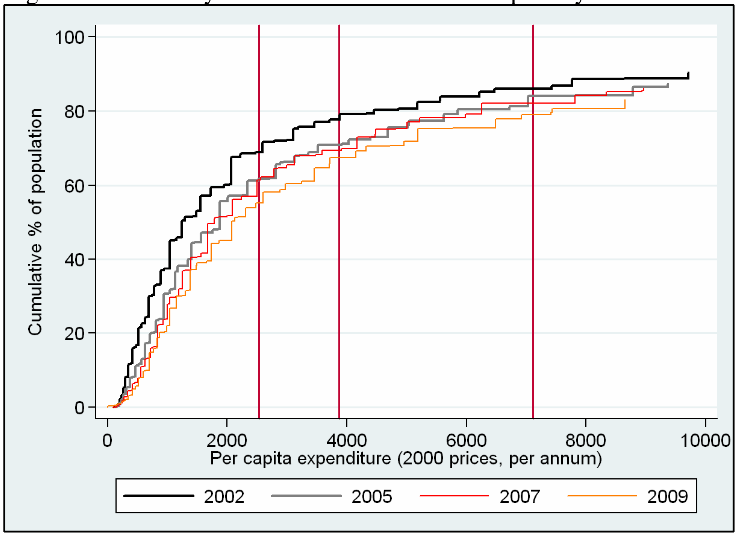


Figure C.13: Poverty headcount ratios at different poverty lines: Selected GHSs (After SRMI2)

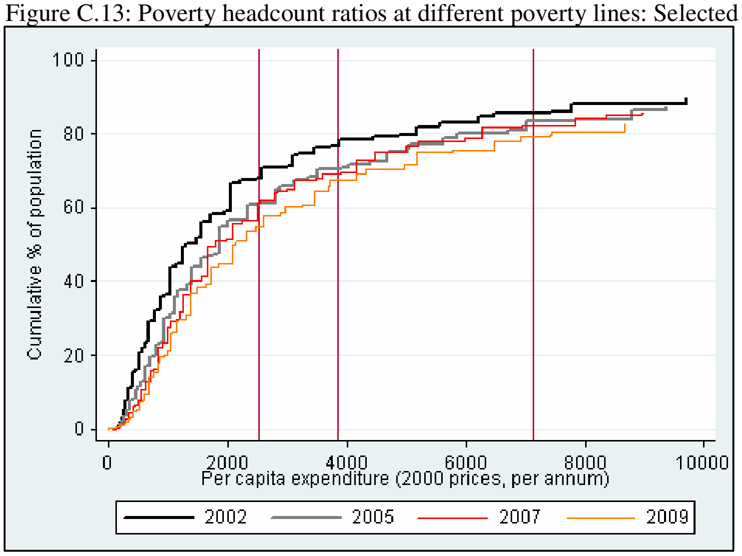


Figure C.14: Poverty headcount ratios at different poverty lines: PSLSD

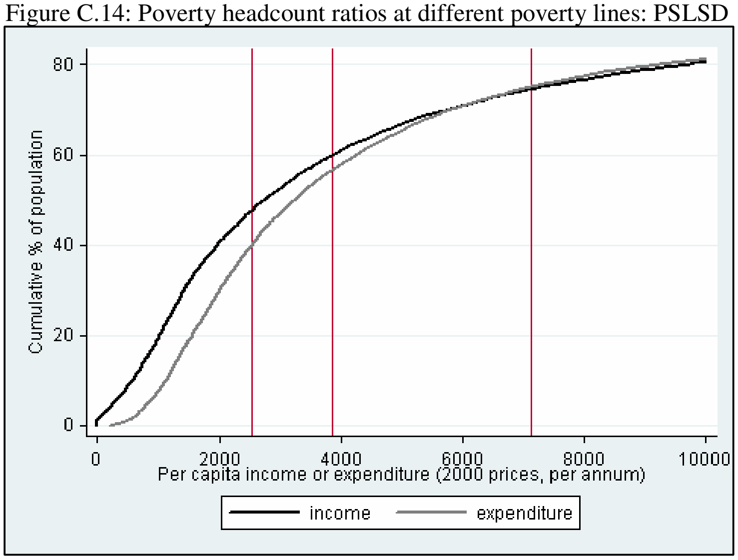


Figure C.15: Poverty headcount ratios at different poverty lines: NIDS

Figure C.16: Poverty headcount ratios at different poverty lines: Selected AMPSs

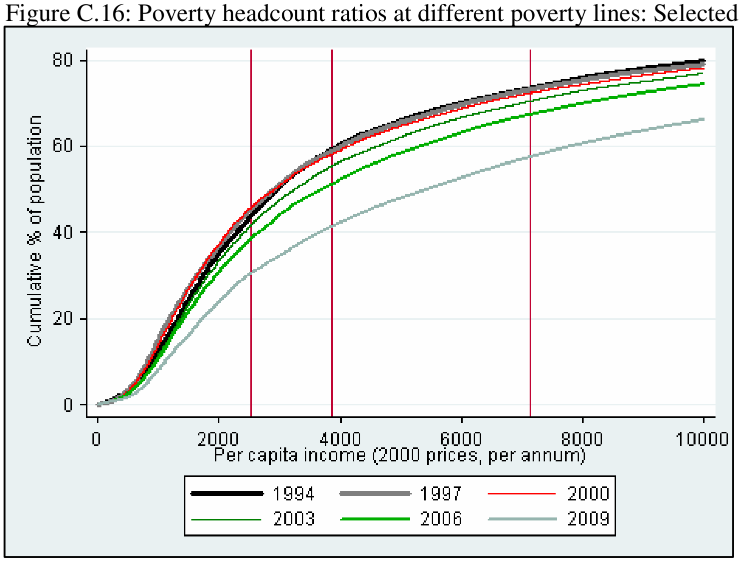


Figure C.17: Poverty gap indices at different poverty lines: Census (No imputations)

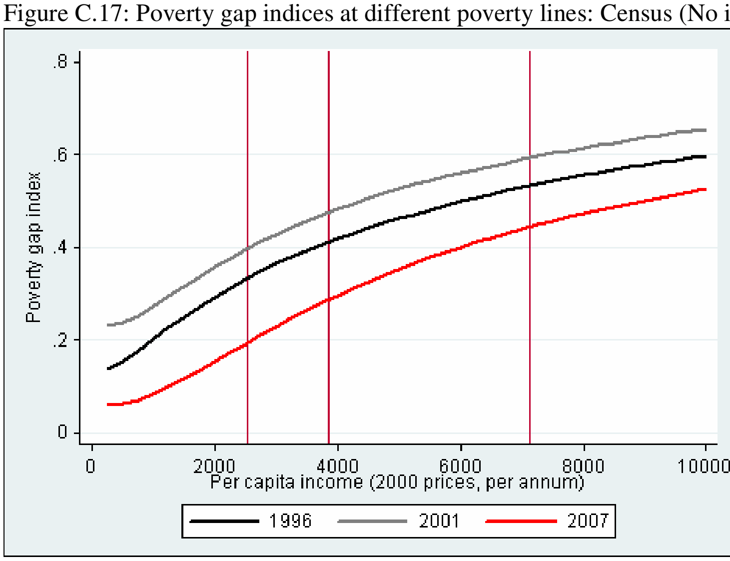


Figure C.18: Poverty gap indices at different poverty lines: Census (After SRMI1)

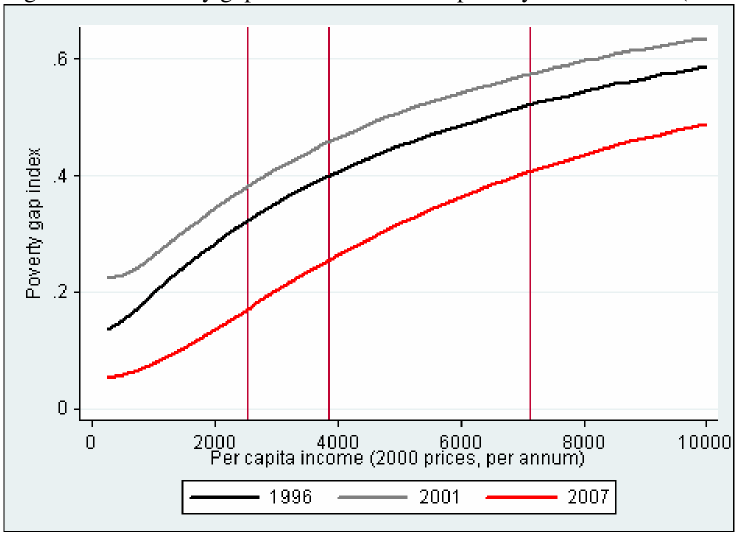


Figure C.19: Poverty gap indices at different poverty lines: Census (After SRMI2)

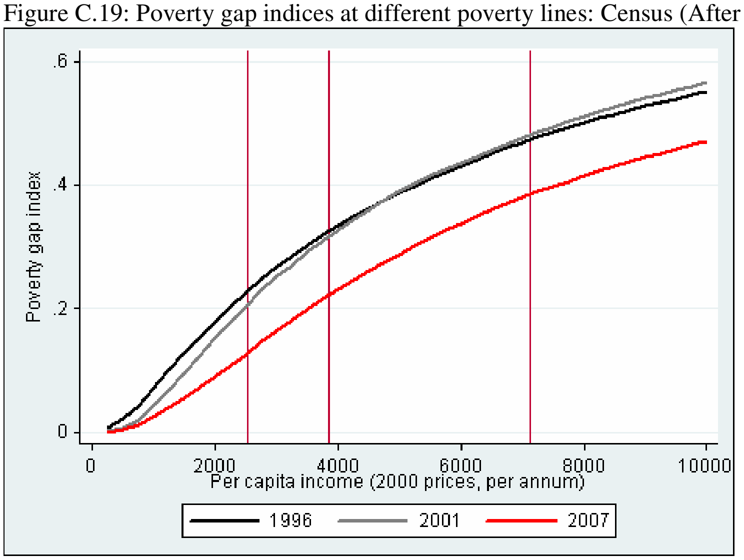


Figure C.20: Poverty gap indices at different poverty lines: IES Income (STC)

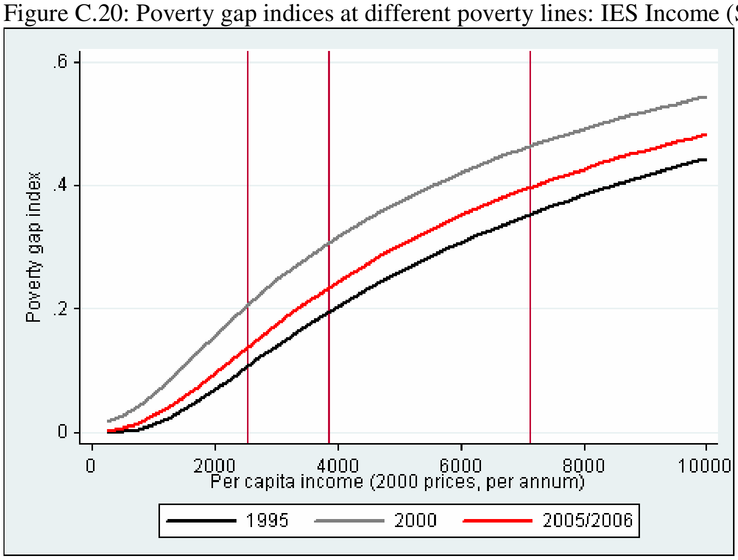


Figure C.21: Poverty gap indices at different poverty lines: IES Expenditure (STC)

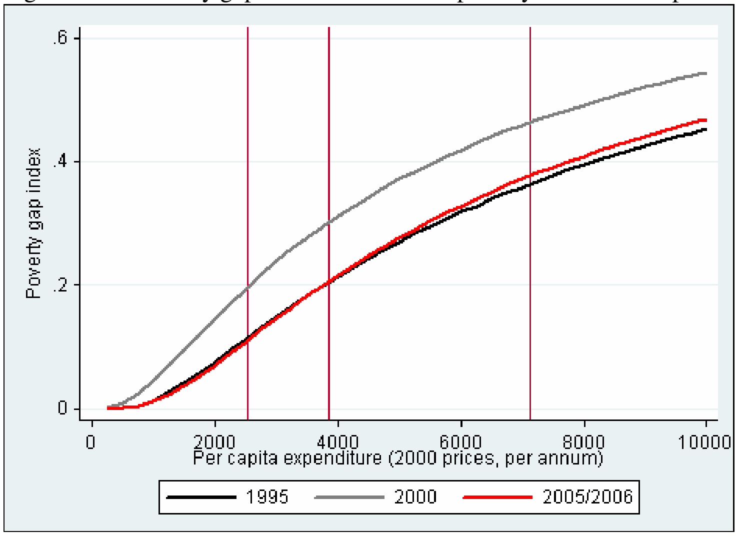


Figure C.22: Poverty gap indices at different poverty lines: IES Income (COICOP)

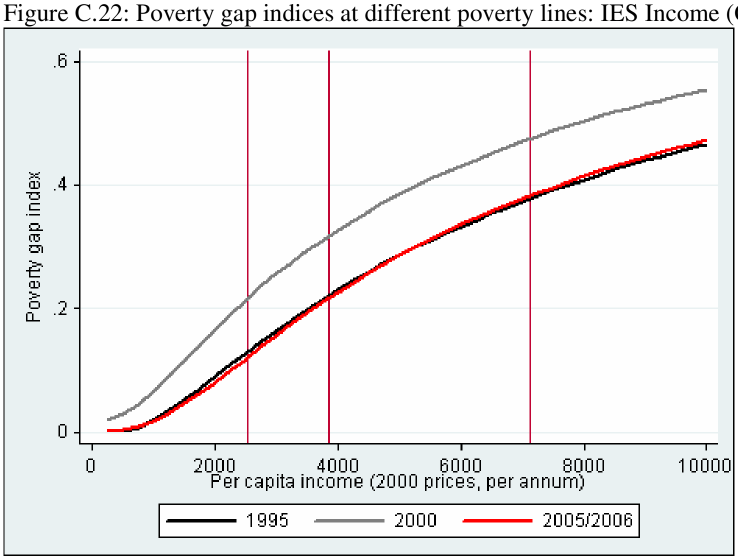


Figure C.23: Poverty gap indices at different poverty lines: IES Consumption (COICOP)

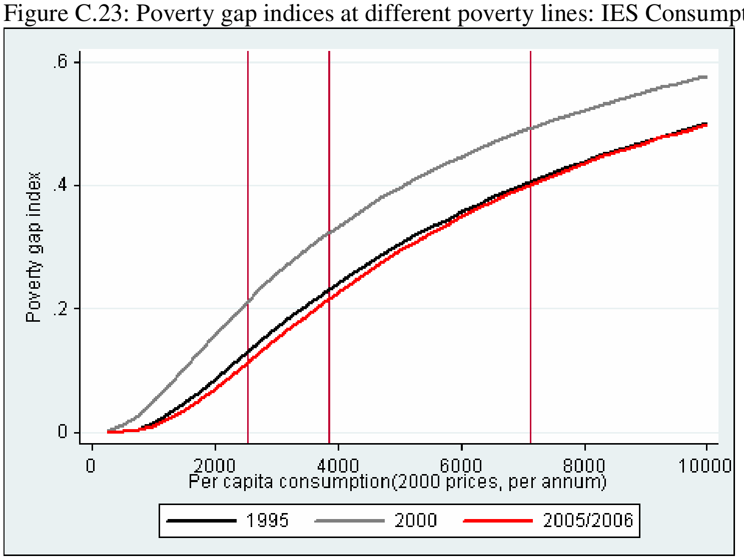


Figure C.24: Poverty gap indices at different poverty lines: Selected OHSs/LFSs (No imputations)

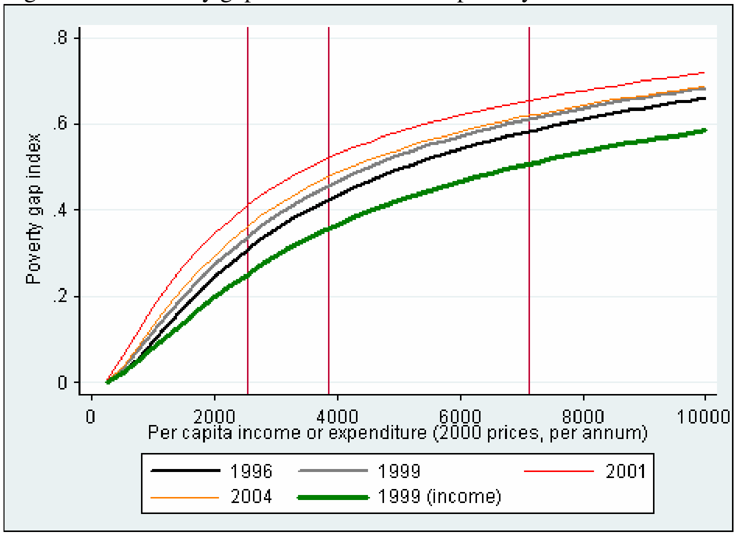


Figure C.25: Poverty gap indices at different poverty lines: Selected OHSs/LFSs (After SRMI2)

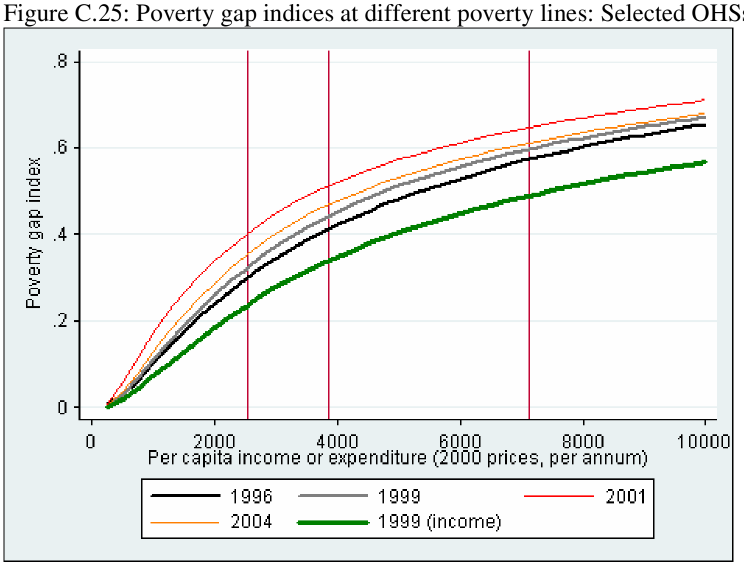


Figure C.26: Poverty gap indices at different poverty lines: Selected GHSs (No imputations)

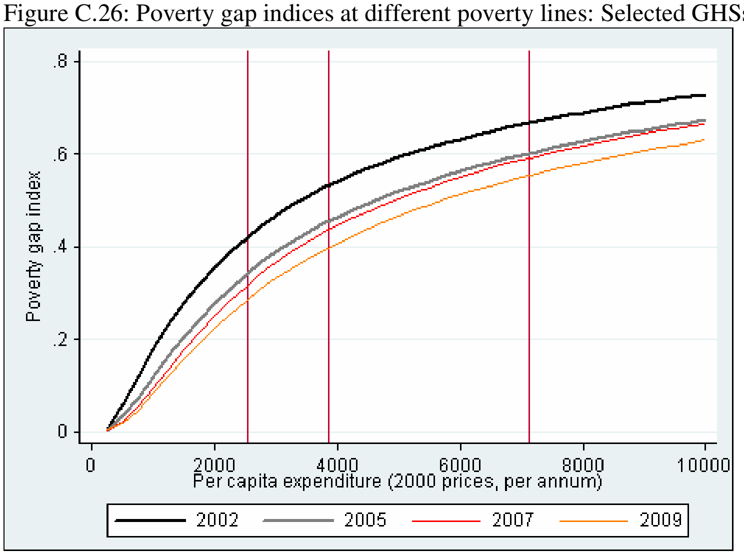


Figure C.27: Poverty gap indices at different poverty lines: Selected GHSs (After SRMI2)

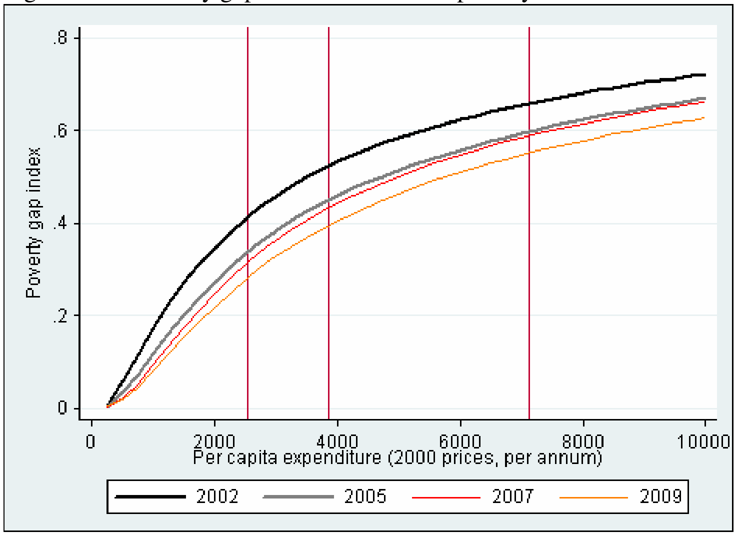


Figure C.28: Poverty gap indices at different poverty lines: PSLSD

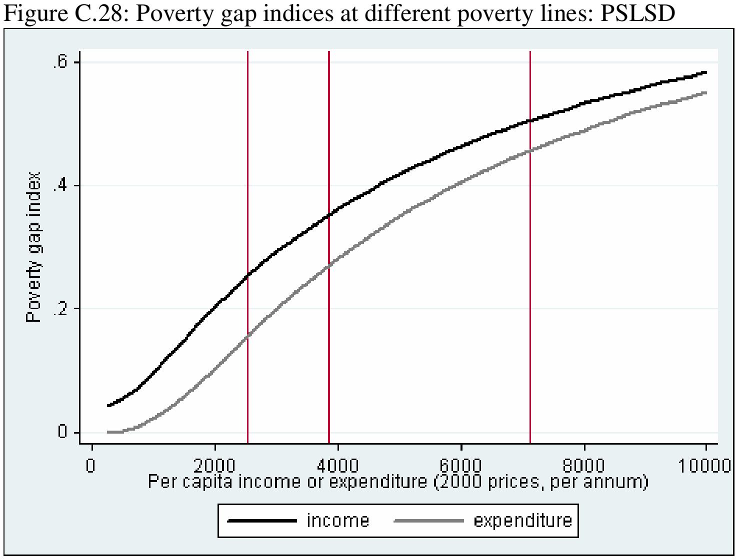


Figure C.29: Poverty gap indices at different poverty lines: NIDS

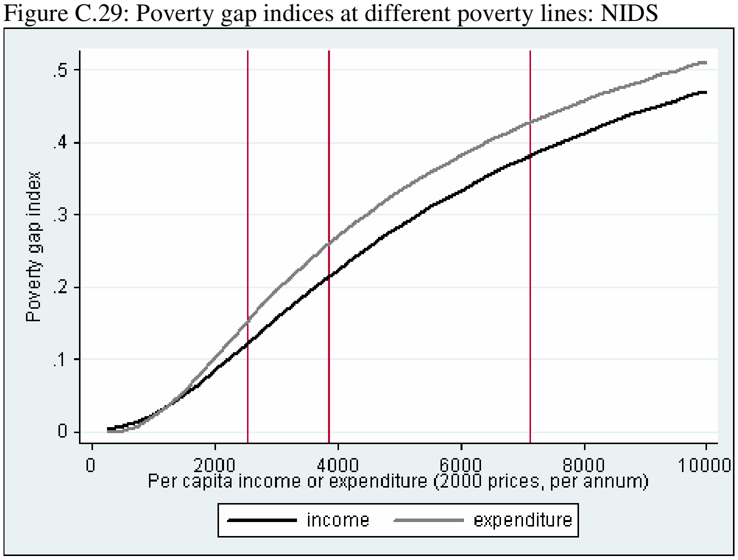


Figure C.30: Poverty gap indices at different poverty lines: Selected AMPSs

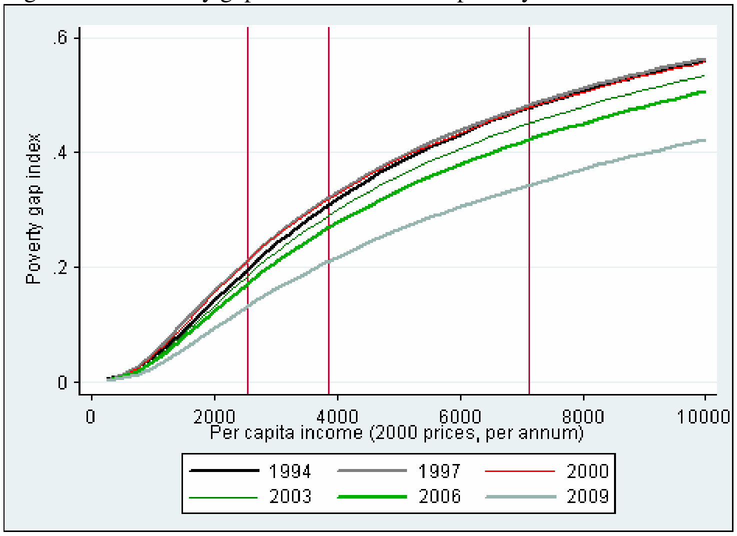


Figure C.31: Squared poverty gap indices at different poverty lines: Census (No imputations)

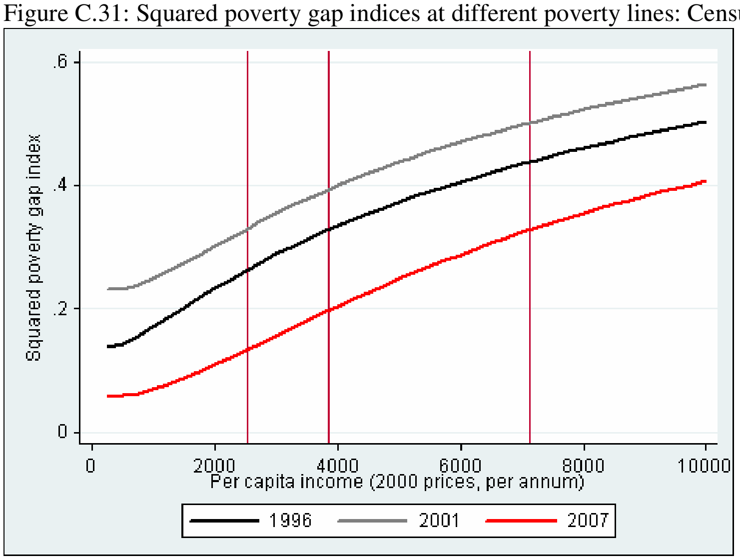


Figure C.32: Squared poverty gap indices at different poverty lines: Census (After SRMI1)

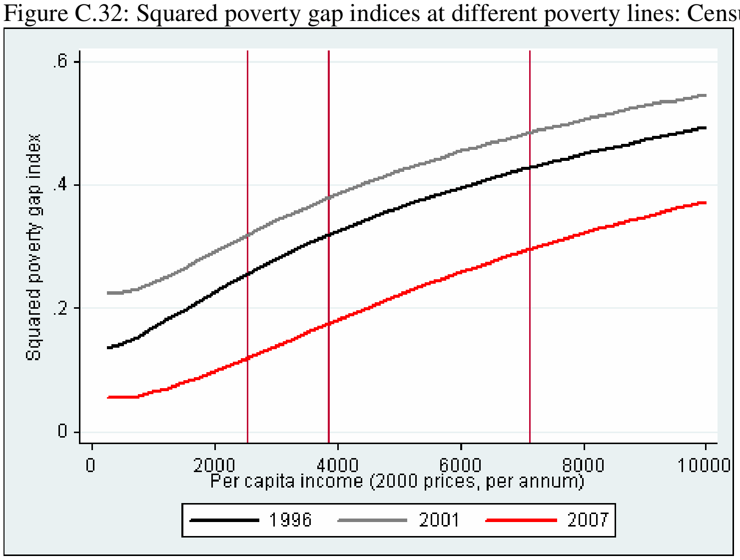


Figure C.33: Squared poverty gap indices at different poverty lines: Census (After SRMI2)

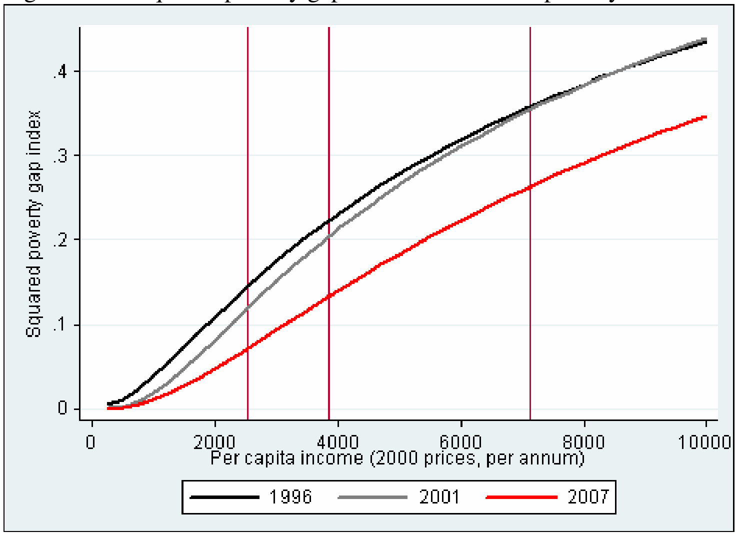


Figure C.34: Squared poverty gap indices at different poverty lines: IES Income (STC)

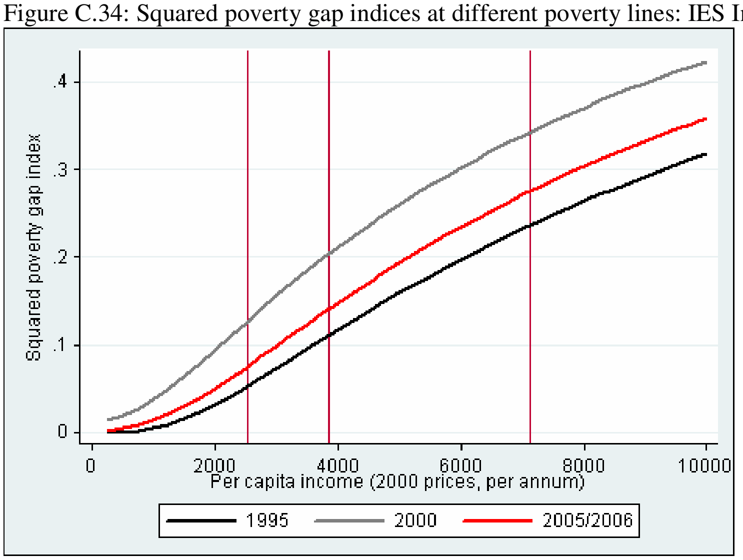


Figure C.35: Squared poverty gap indices at different poverty lines: IES Expenditure (STC)

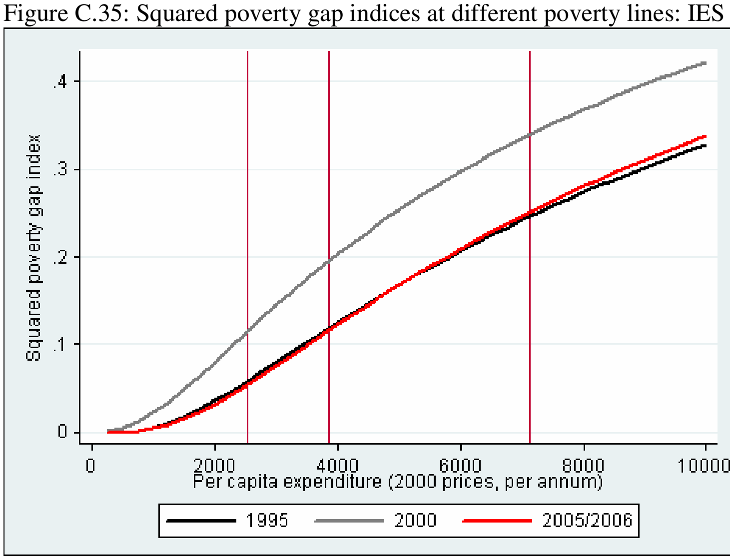


Figure C.36: Squared poverty gap indices at different poverty lines: IES Income (COICOP)

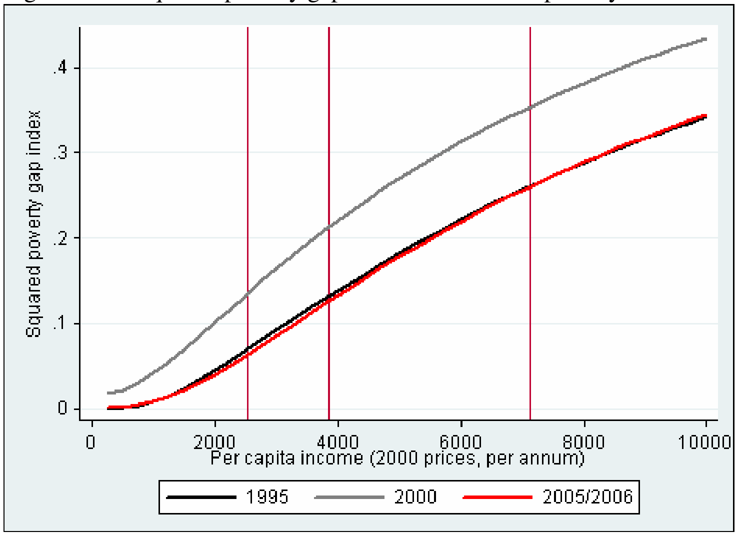


Figure C.37: Squared poverty gap indices at different poverty lines: IES Consumption (COICOP)

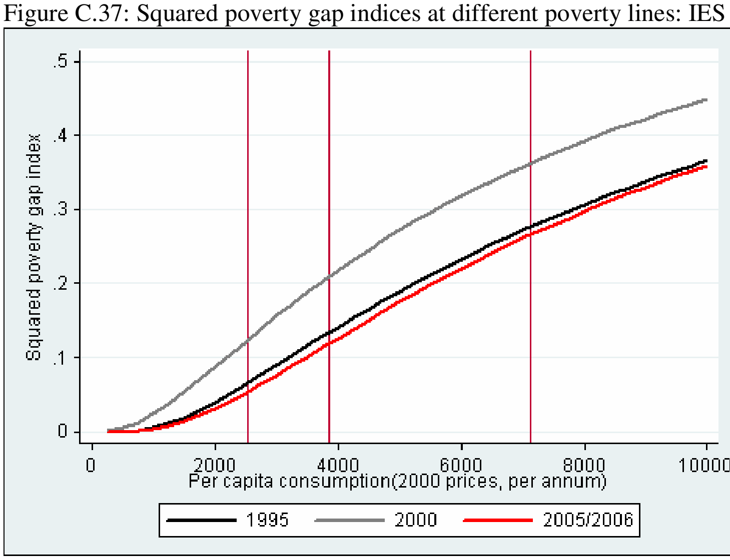


Figure C.38: Squared poverty gap indices at different poverty lines: Selected OHSs/LFSs (No imputations)

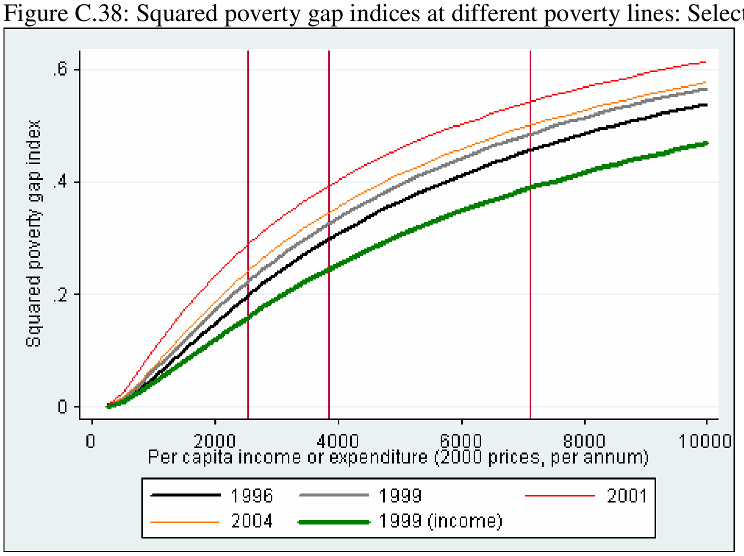


Figure C.39: Squared poverty gap indices at different poverty lines: Selected OHSs/LFSs (After SRMI2)

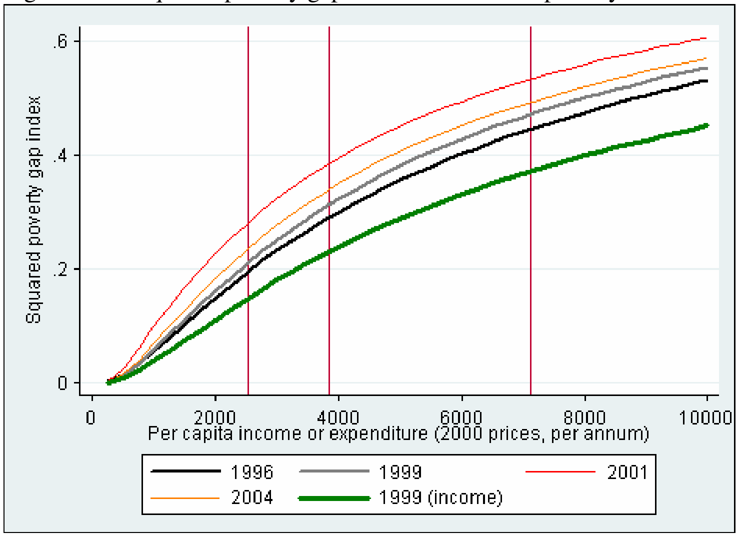


Figure C.40: Squared poverty gap indices at different poverty lines: Selected GHSs (No imputations)

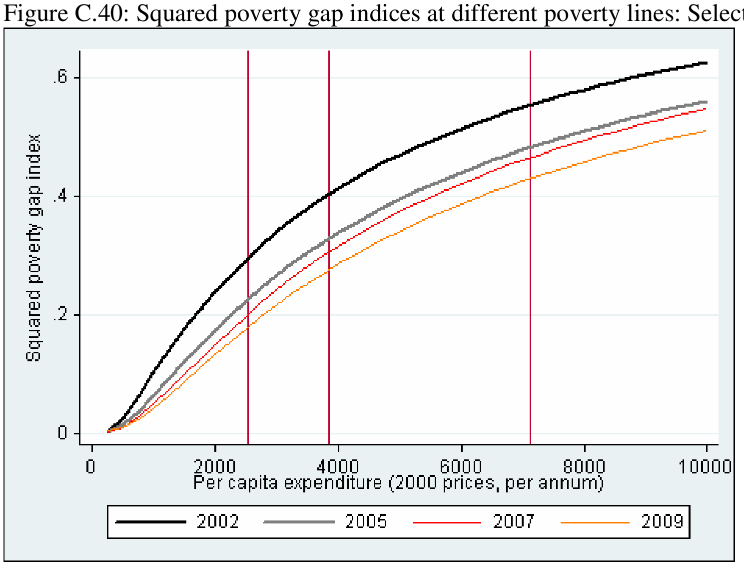


Figure C.41: Squared poverty gap indices at different poverty lines: Selected GHSs (After SRMI2)

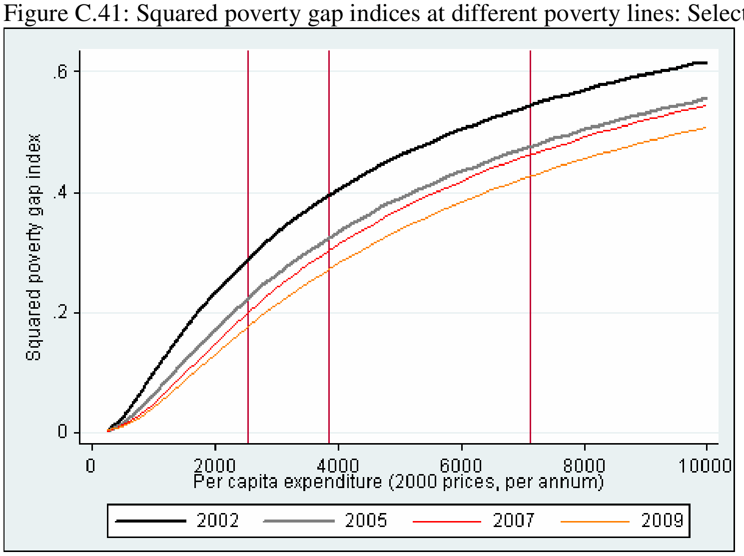


Figure C.42: Squared poverty gap indices at different poverty lines: PSLSD

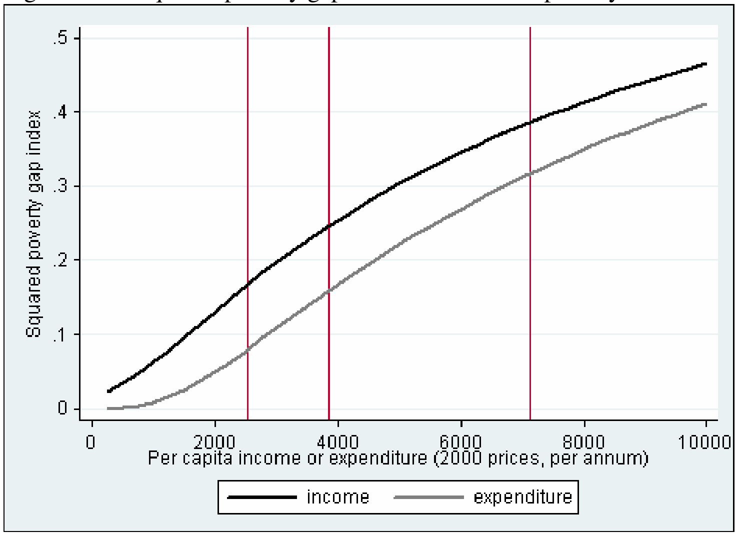


Figure C.43: Squared poverty gap indices at different poverty lines: NIDS

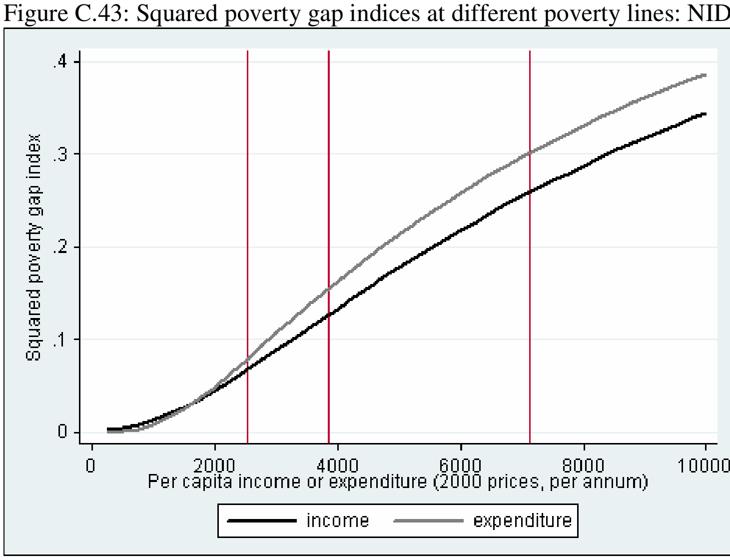


Figure C.44: Squared poverty gap indices at different poverty lines: Selected AMPSs

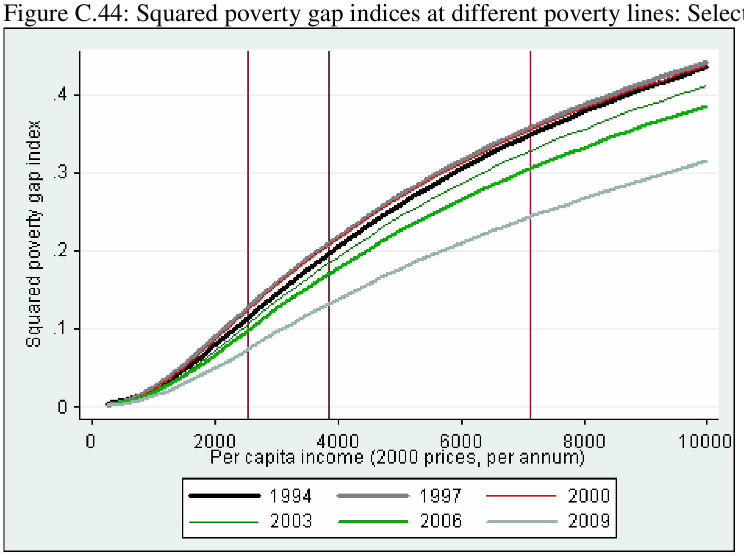


Figure C.45: Lorenz curves: Census (No imputations)

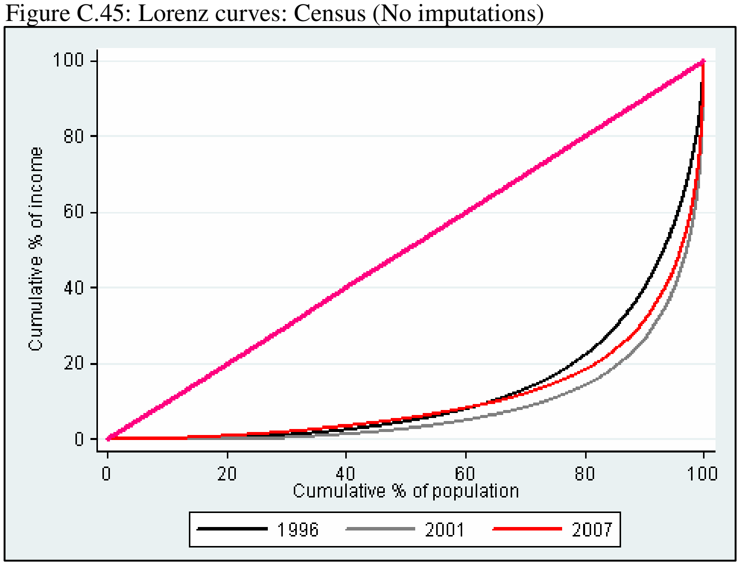


Figure C.46: Lorenz curves: Census (After SRMI1)

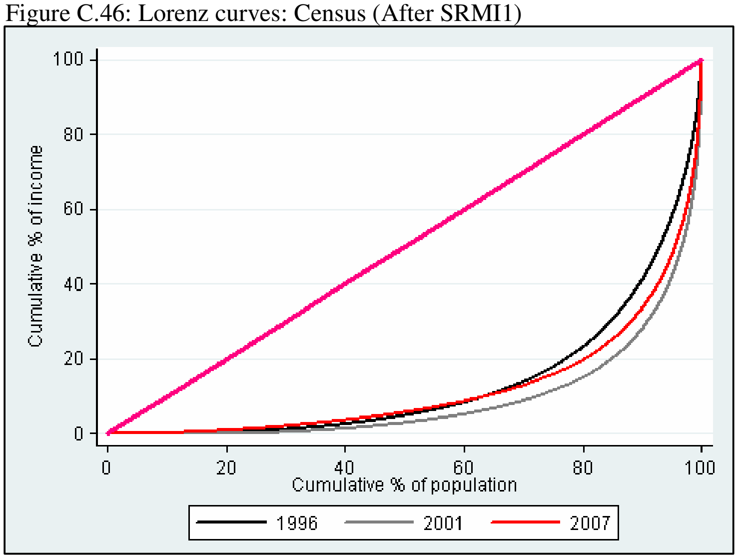


Figure C.47: Lorenz curves: Census (After SRMI2)

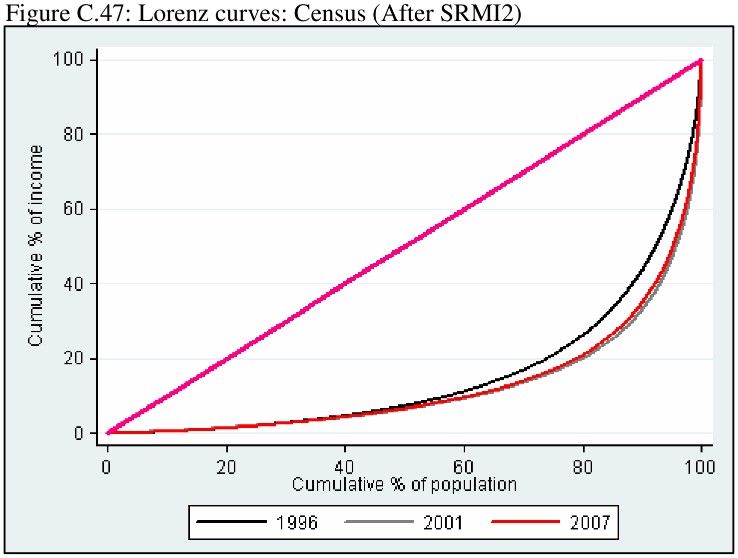


Figure C.48: Lorenz curves: IES Income (STC)

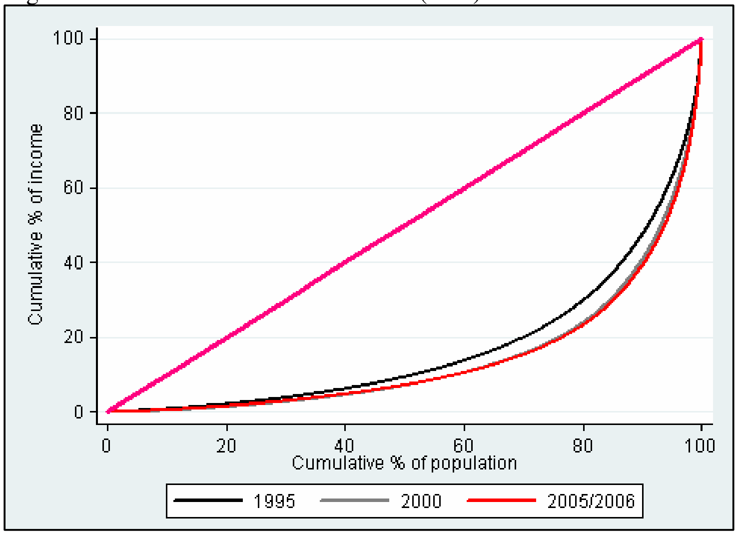


Figure C.49: Lorenz curves: IES Expenditure (STC)

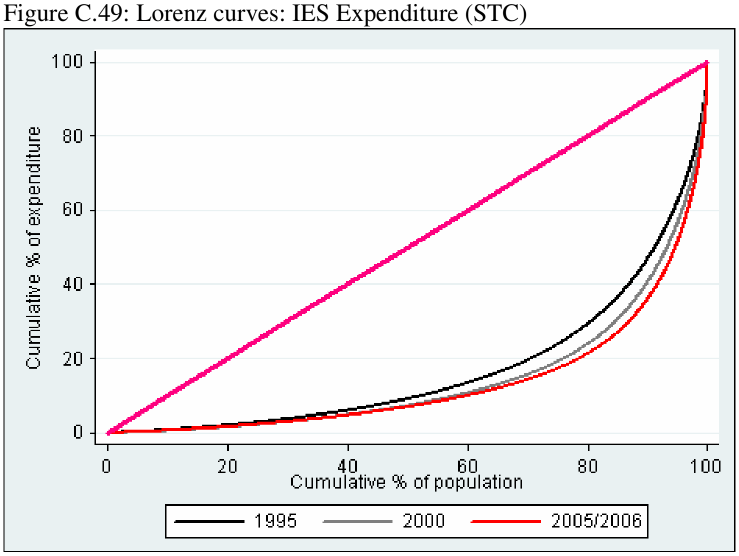


Figure C.50: Lorenz curves: IES Income (COICOP)

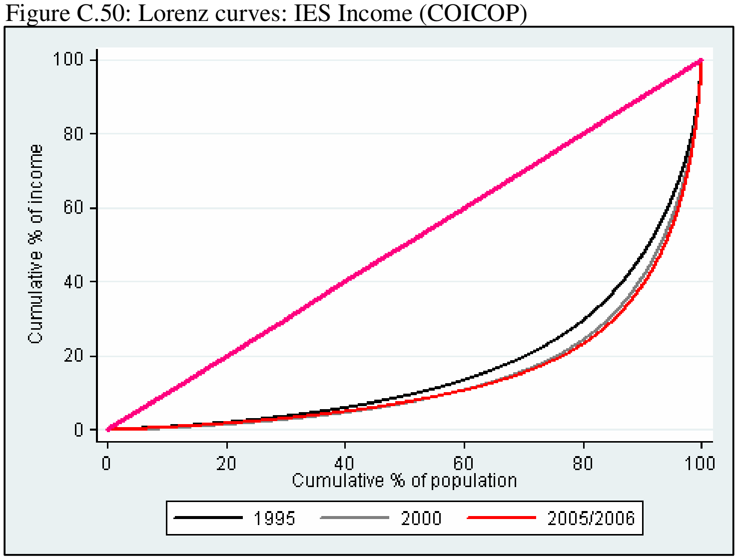


Figure C.51: Lorenz curves: IES Consumption (COICOP)

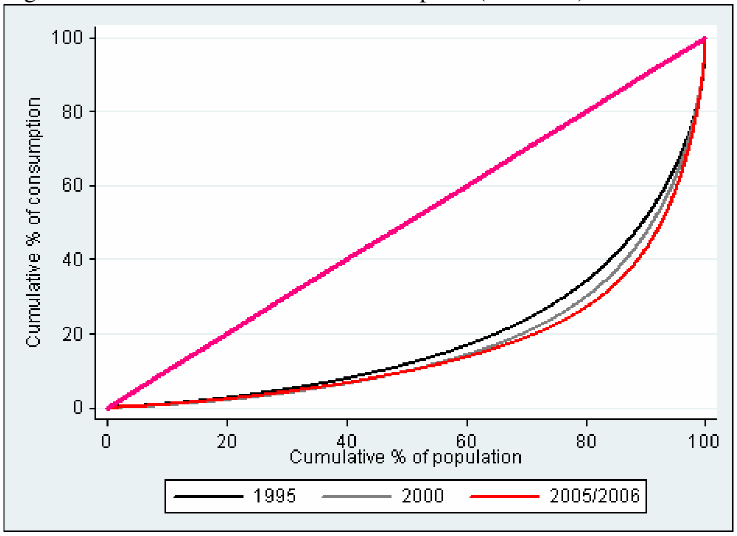


Figure C.52: Lorenz curves: Selected OHSs/LFSs (No imputations)

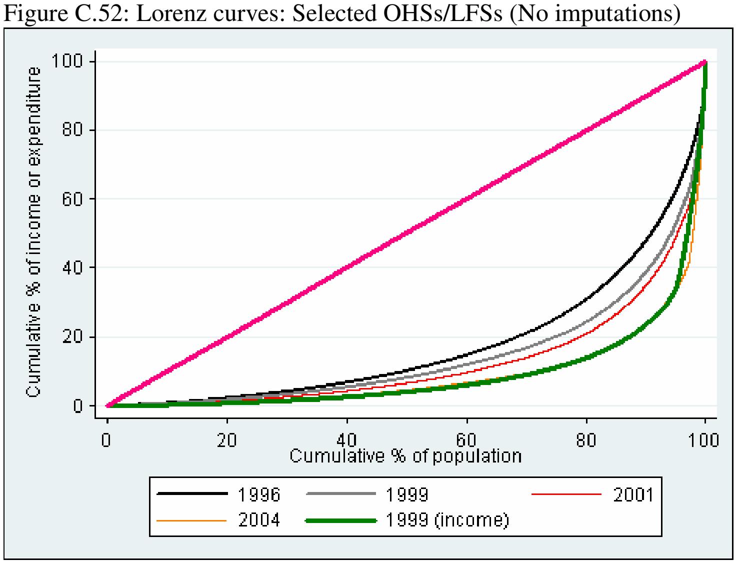


Figure C.53: Lorenz curves: Selected OHSs/LFSs (After SRMI2)

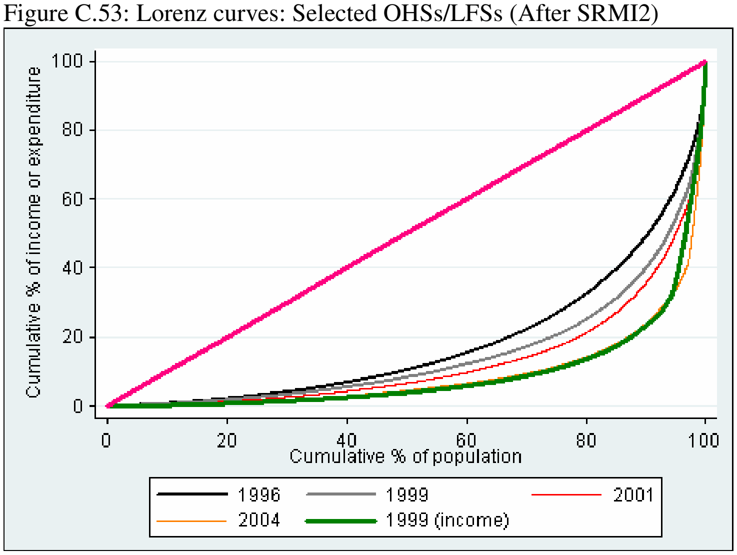


Figure C.54: Lorenz curves: Selected GHSs (No imputations)

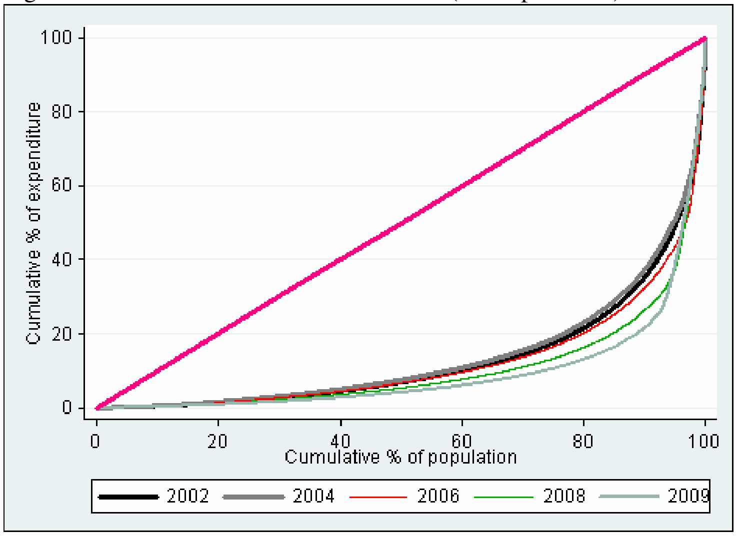


Figure C.55: Lorenz curves: Selected GHSs (After SRMI2)

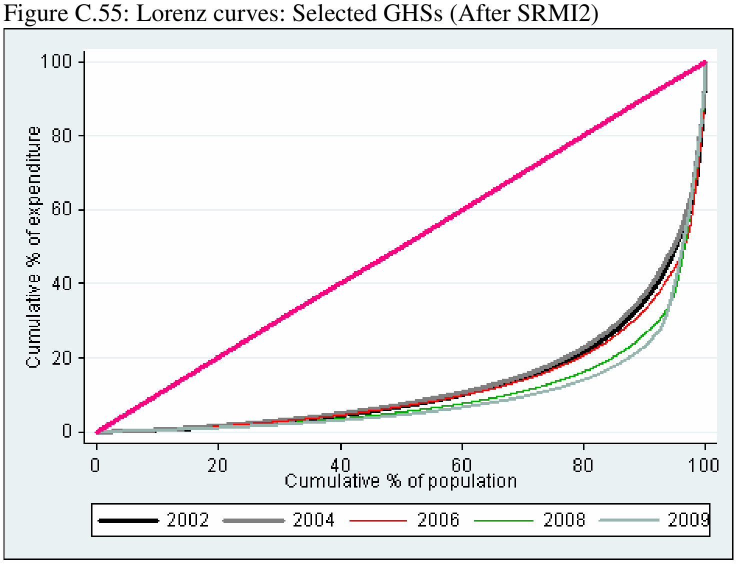


Figure C.56: Lorenz curves: PSLSD

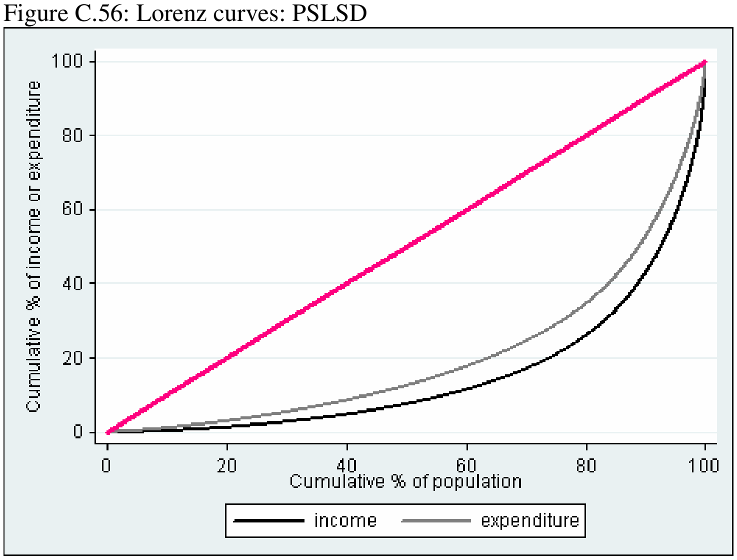


Figure C.57: Lorenz curves: NIDS

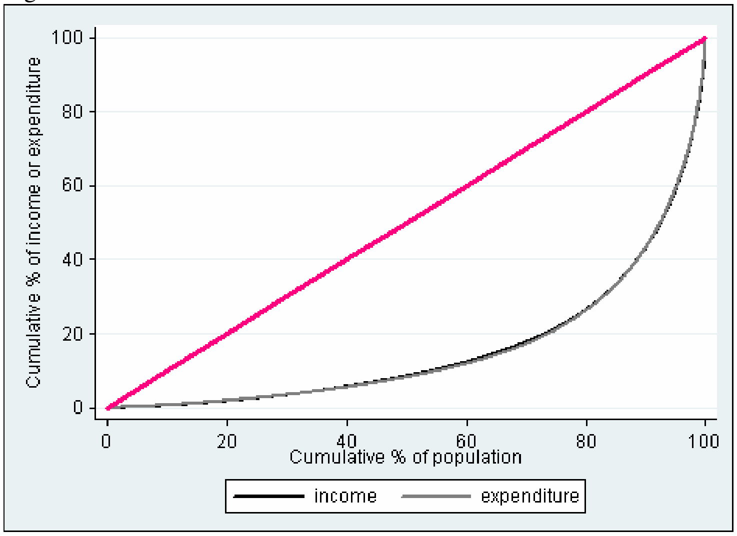


Figure C.58: Lorenz curves: Selected AMPSs

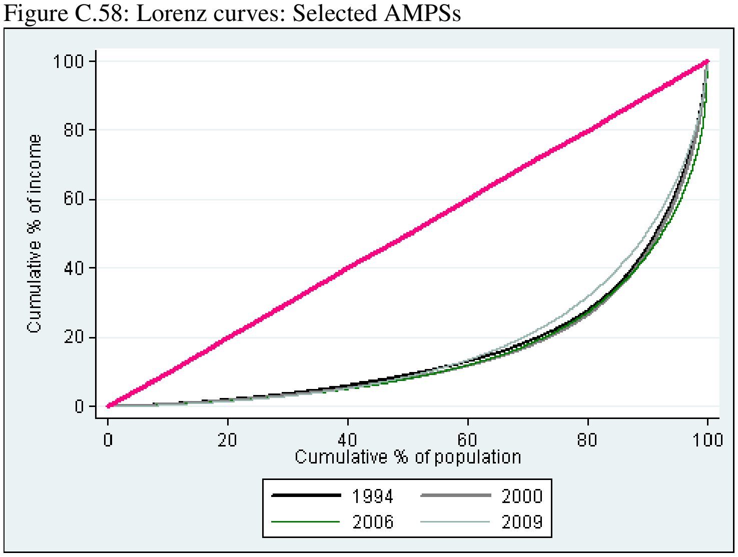


Figure C.59: Within-race and between-race's shares of total inequality in the Theil-L index in AMPSs

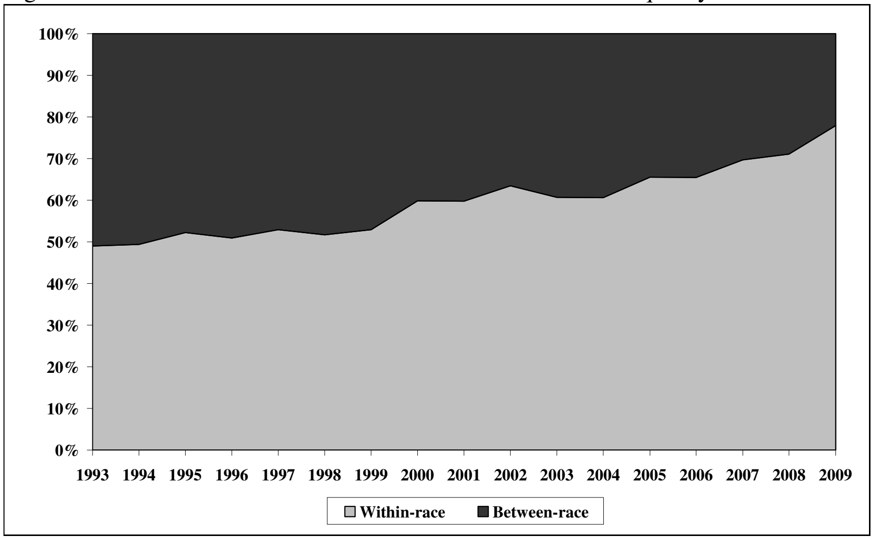


Figure C.60: Within-race and between-race's shares of total inequality in the Theil-L index in Censuses and CS 2007

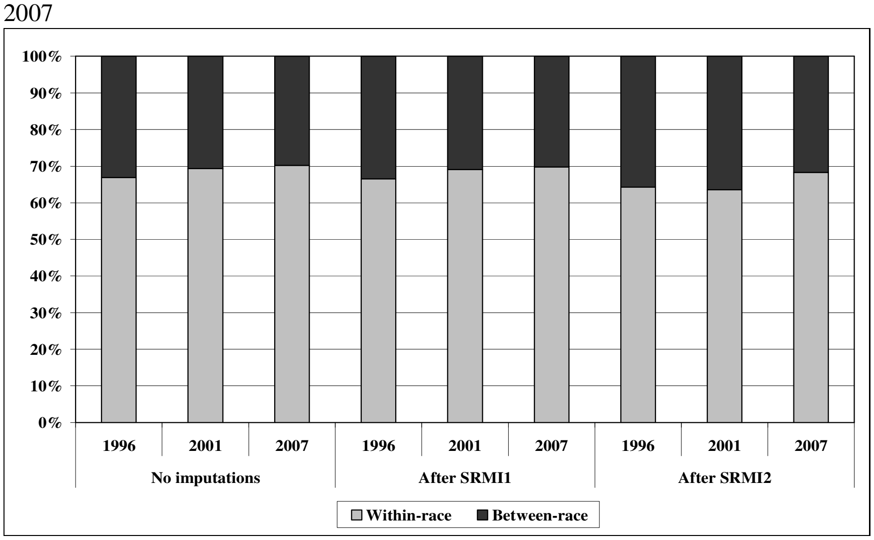


Figure C.61: Within-race and between-race's shares of total inequality in the Theil-L index in IESs

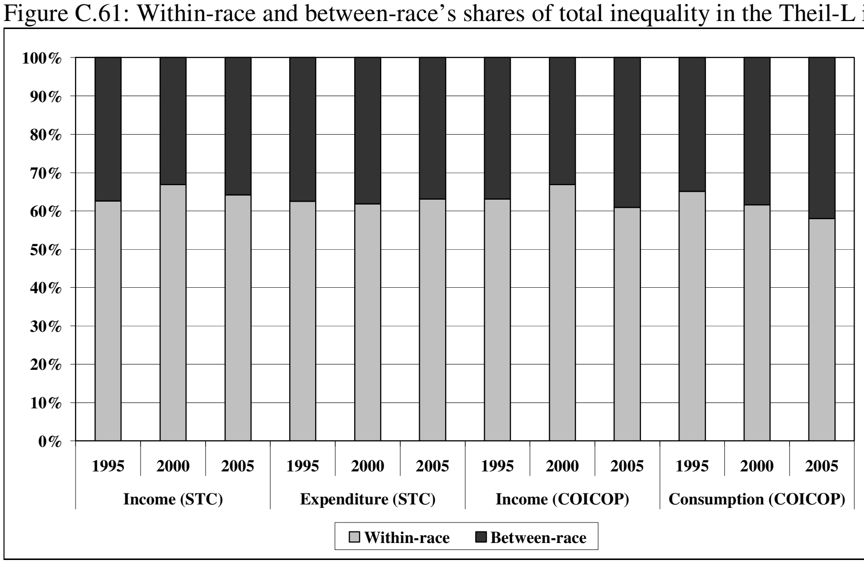


Figure C.62: Within-race and between-race's shares of total inequality in the Theil-T index in AMPSs

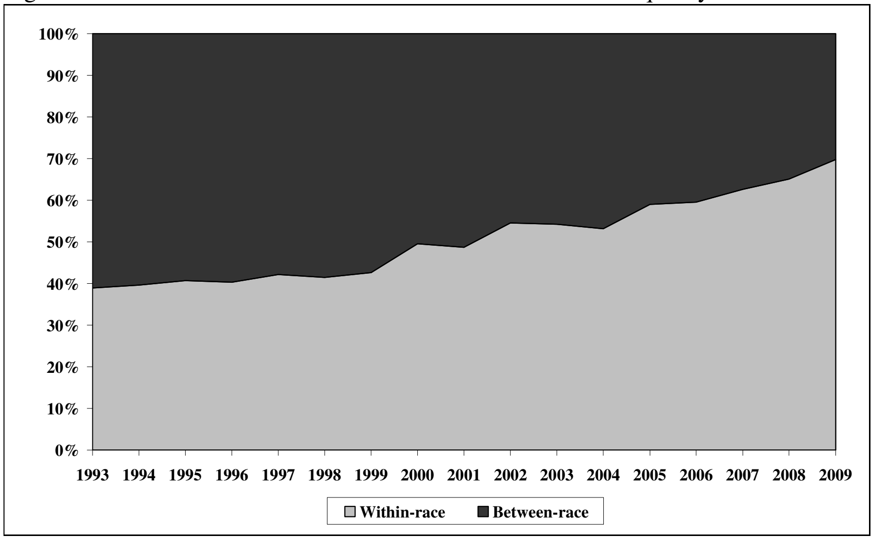


Figure C.63: Within-race and between-race's shares of total inequality in the Theil-T index in Censuses and CS 2007

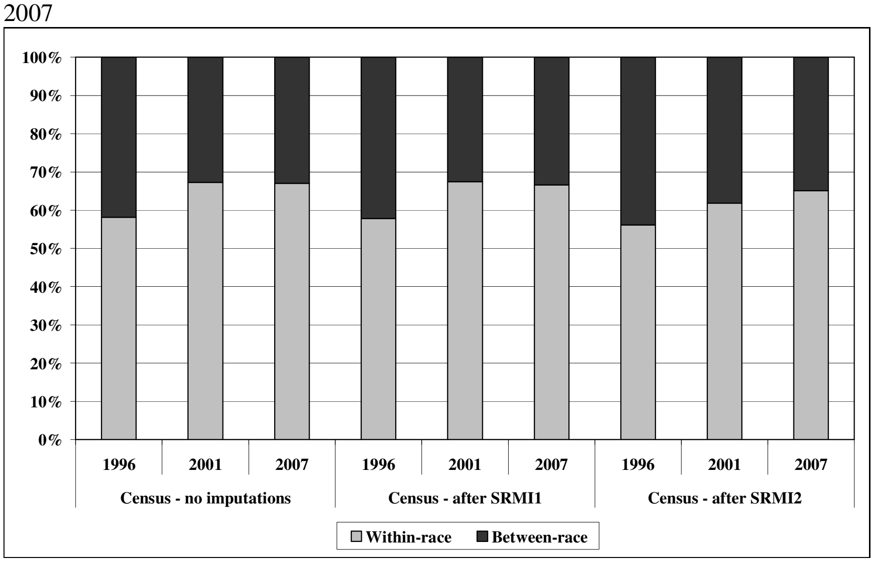


Figure C.64: Within-race and between-race's shares of total inequality in the Theil-T index in IESs

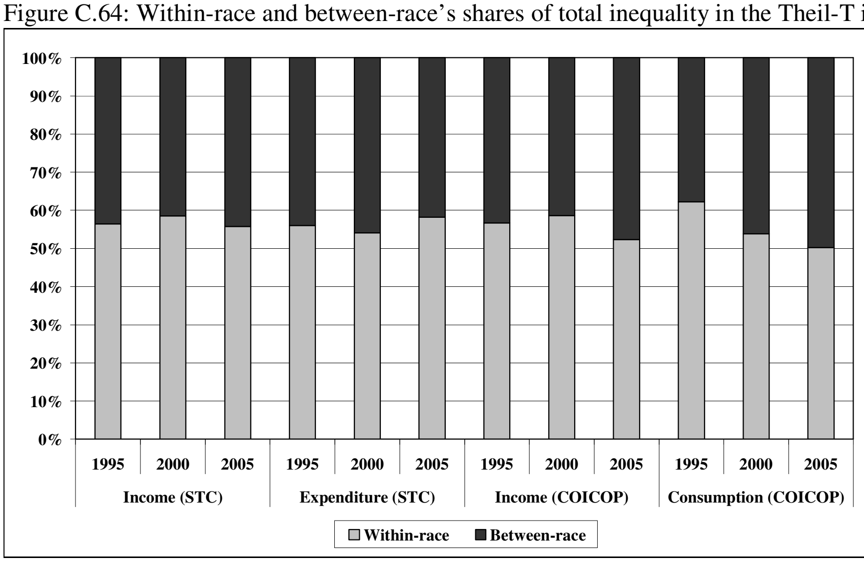


Figure C.65: Growth incidence curve for real per capita income (STC), IES 1995 and IES 2000

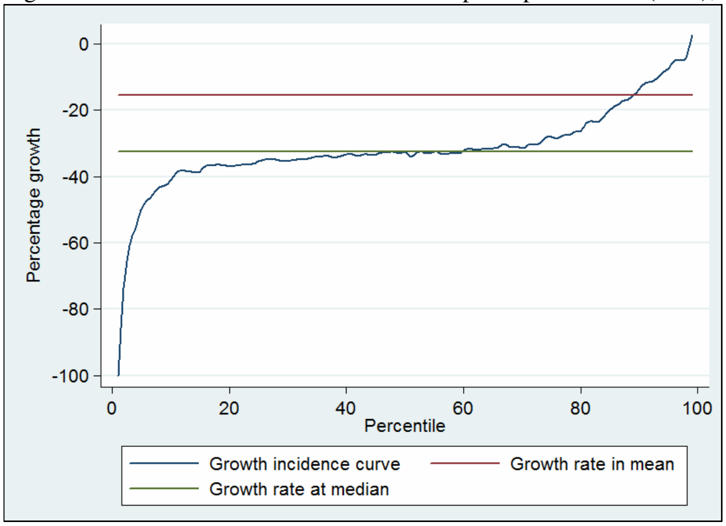


Figure C.66: Growth incidence curve for real per capita expenditure (STC), IES 2000 and IES 2005/2006

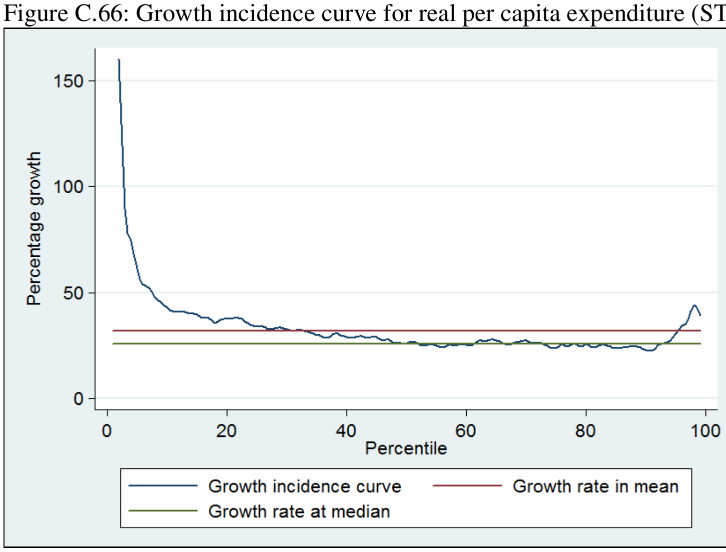


Figure C.67: Growth incidence curve for real per capita expenditure (STC), IES 1995 and IES 2000

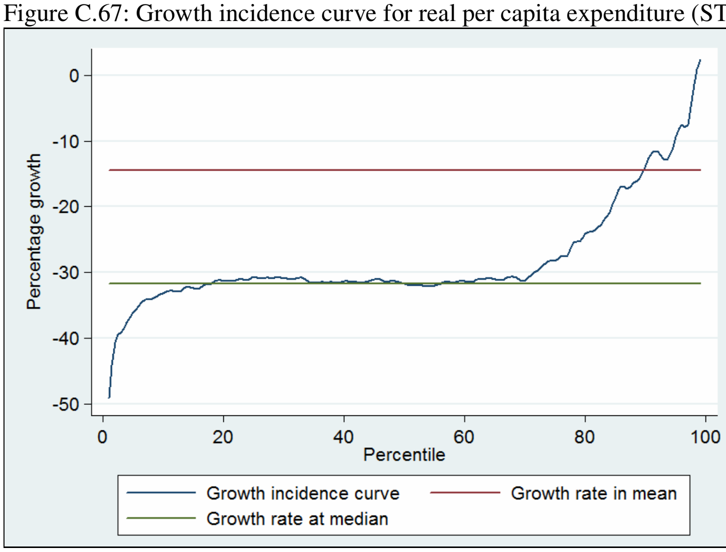


Figure C.68: Growth incidence curve for real per capita expenditure (STC), IES 2000 and IES 2005/2006

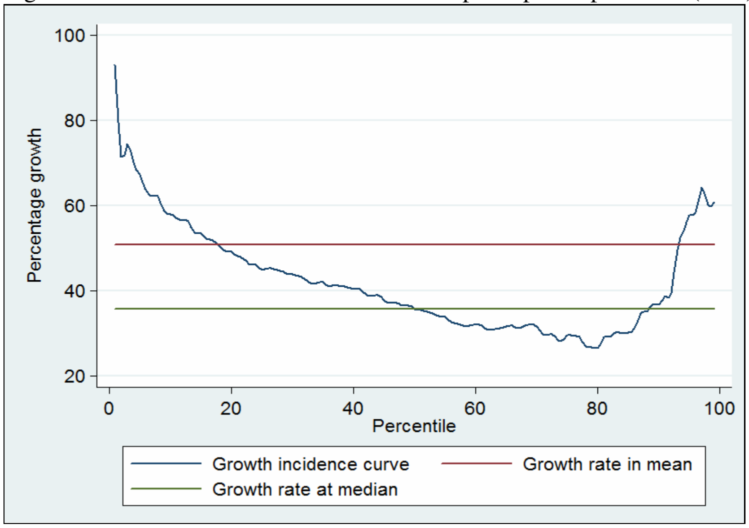


Figure C.69: Growth incidence curve for real per capita income (COICOP), IES 1995 and IES 2000

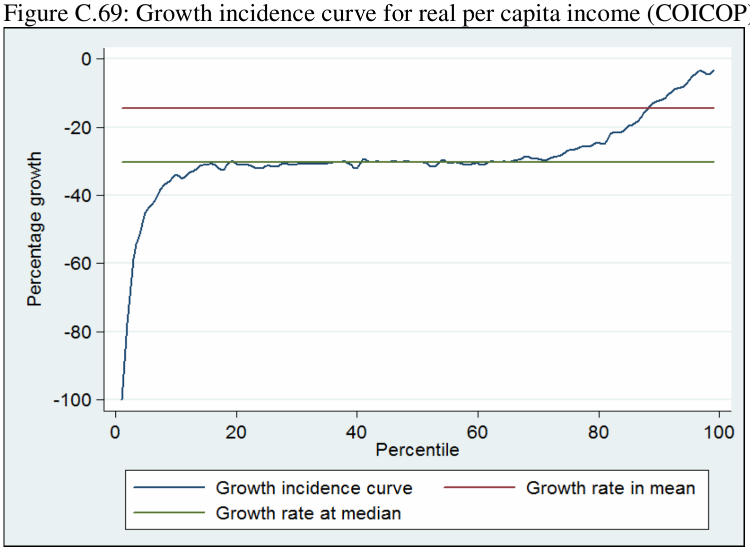


Figure C.70: Growth incidence curve for real per capita income (COICOP), IES 2000 and IES 2005/2006

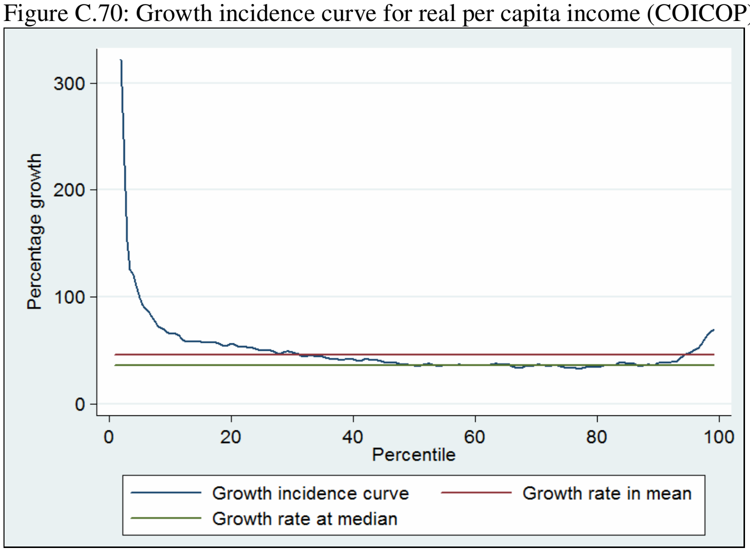


Figure C.71: Growth incidence curve for real per capita consumption (COICOP), IES 1995 and IES 2000

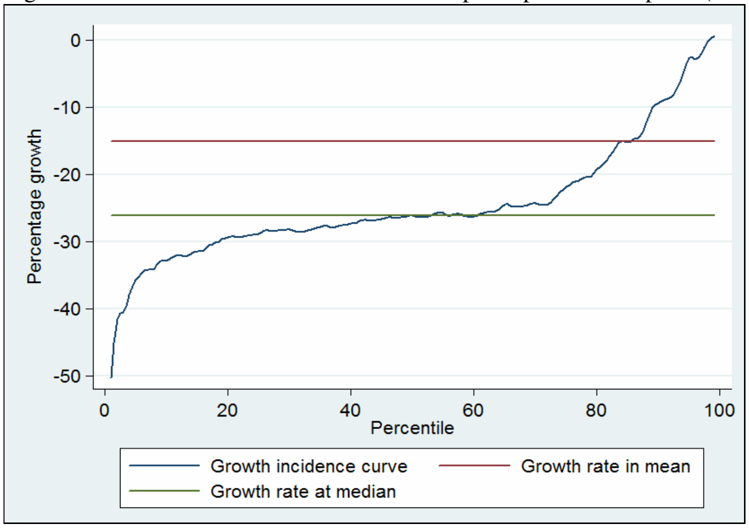


Figure C.72: Growth incidence curve for real per capita consumption (COICOP), IES 2000 and IES 2005/2006

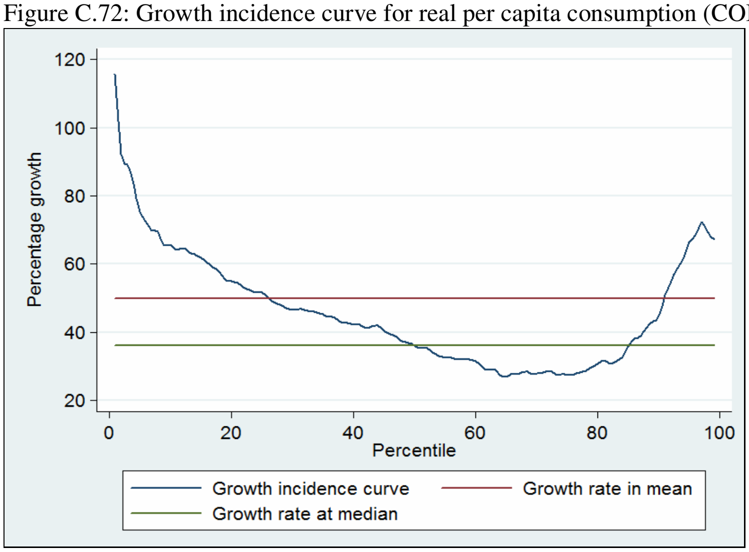


Figure C.73: Growth incidence curve for real per capita income after SRMI2 was applied, Census 1996 and Census 2001

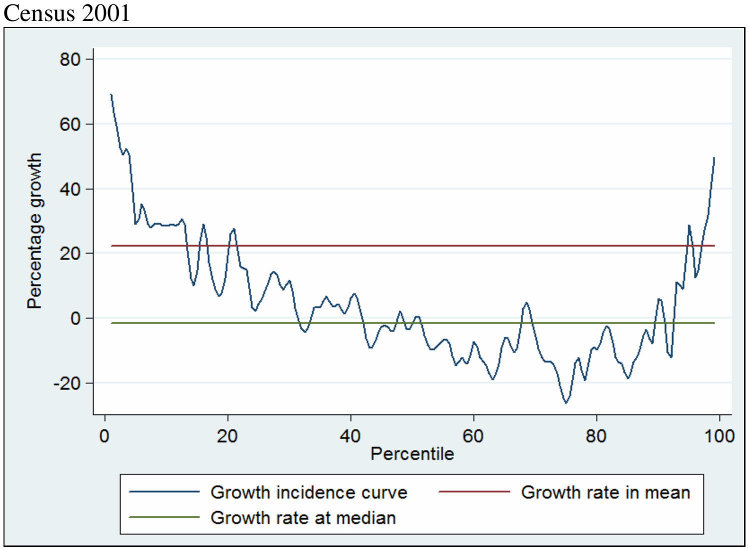


Figure C.74: Growth incidence curve for real per capita income after SRMI2 was applied, Census 2001 and CS 2007

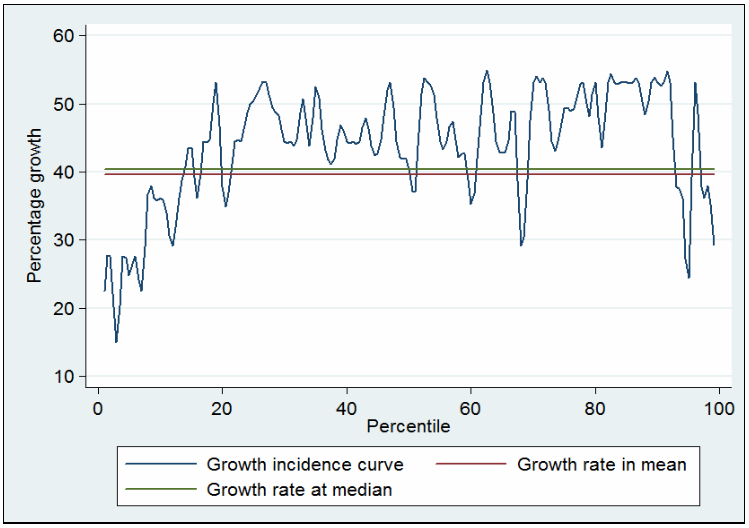


Figure C.75: Growth incidence curve for real per capita income, AMPS 1994 and AMPS 2004

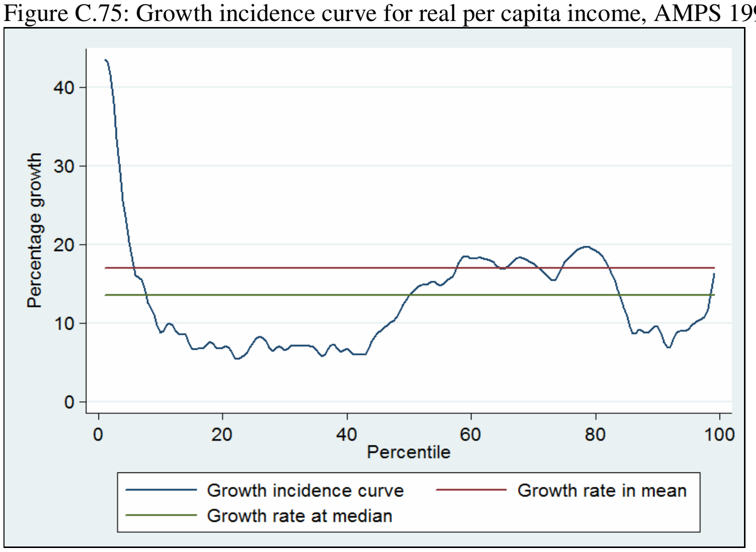
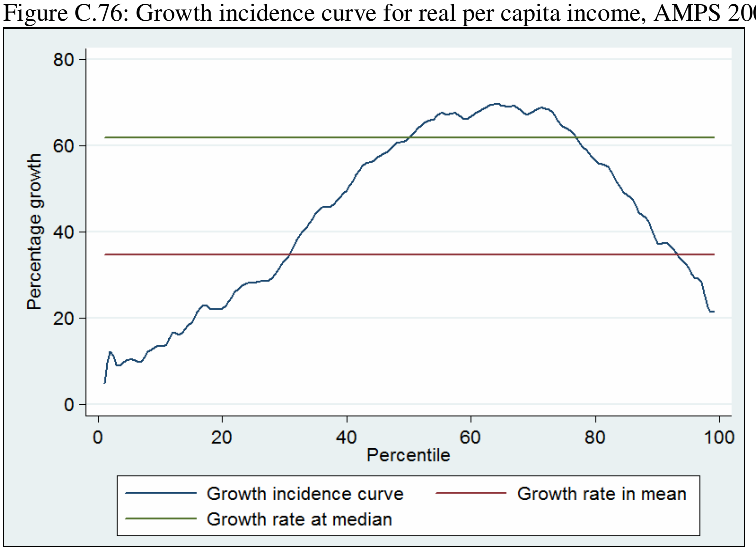


Figure C.76: Growth incidence curve for real per capita income, AMPS 2004 and AMPS 2009



Appendix D: Poverty and inequality estimates in NIDS

Figure D.1: Kernel density curves of log per capita income (2000 prices), NIDS

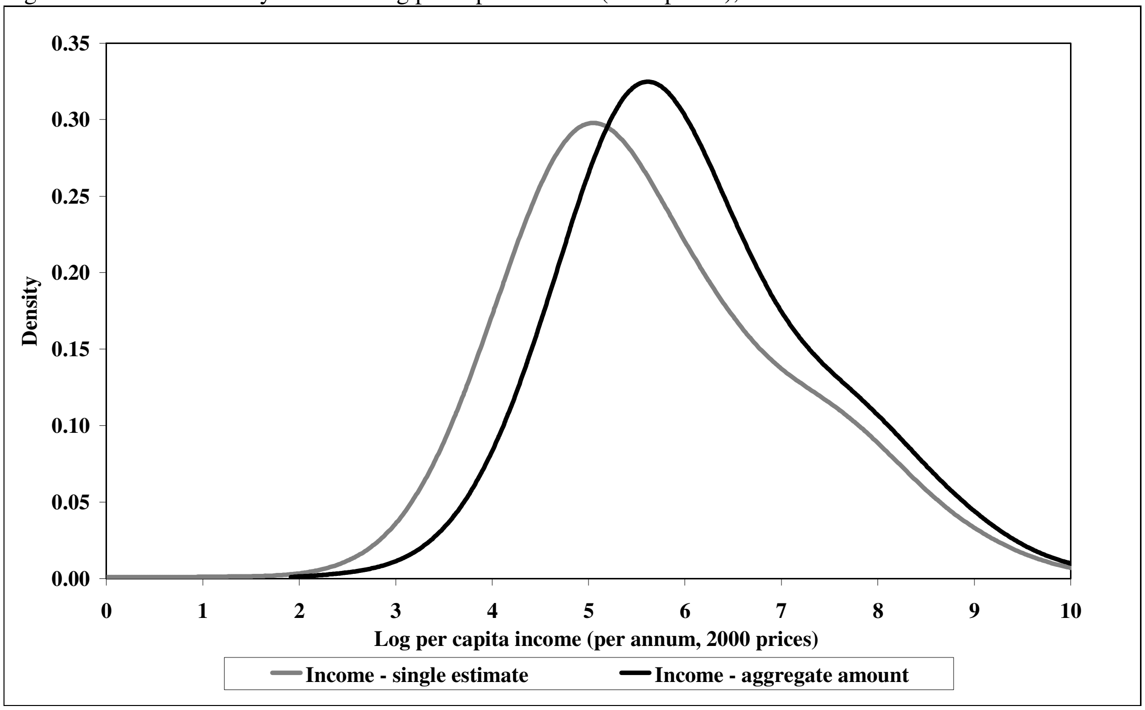


Figure D.2: Kernel density curves of log per capita expenditure (2000 prices), NIDS

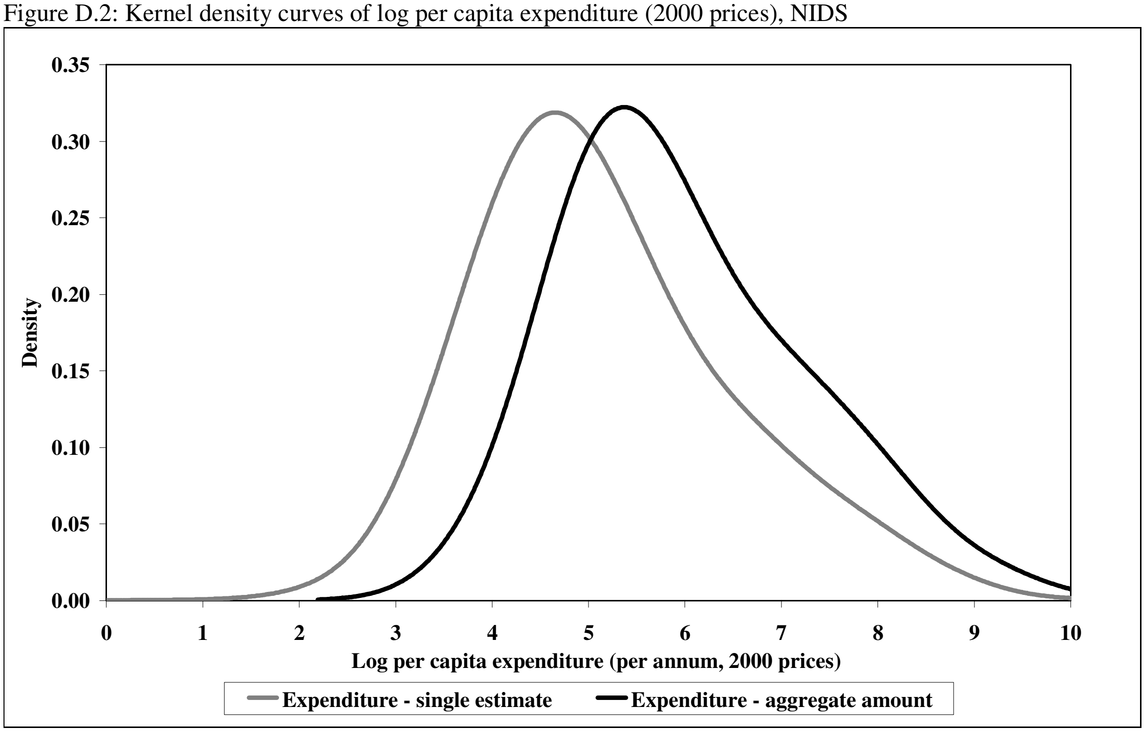


Figure D.3: Poverty headcount ratios at different poverty lines: NIDS income

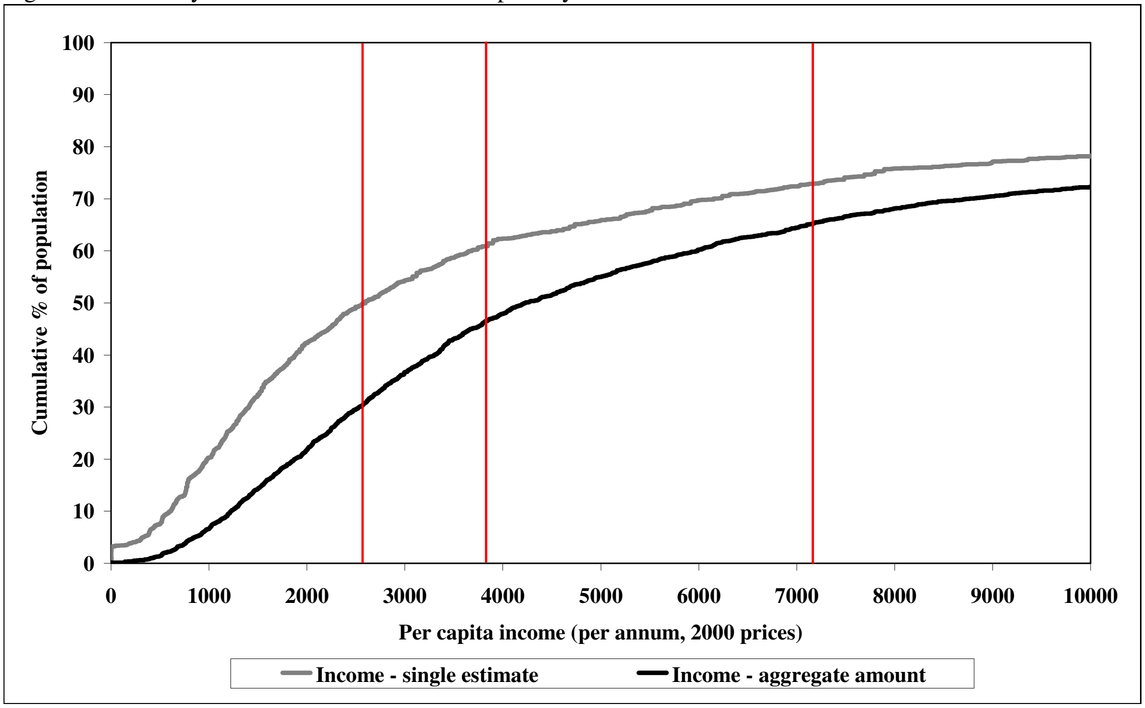


Figure D.4: Poverty headcount ratios at different poverty lines: NIDS expenditure

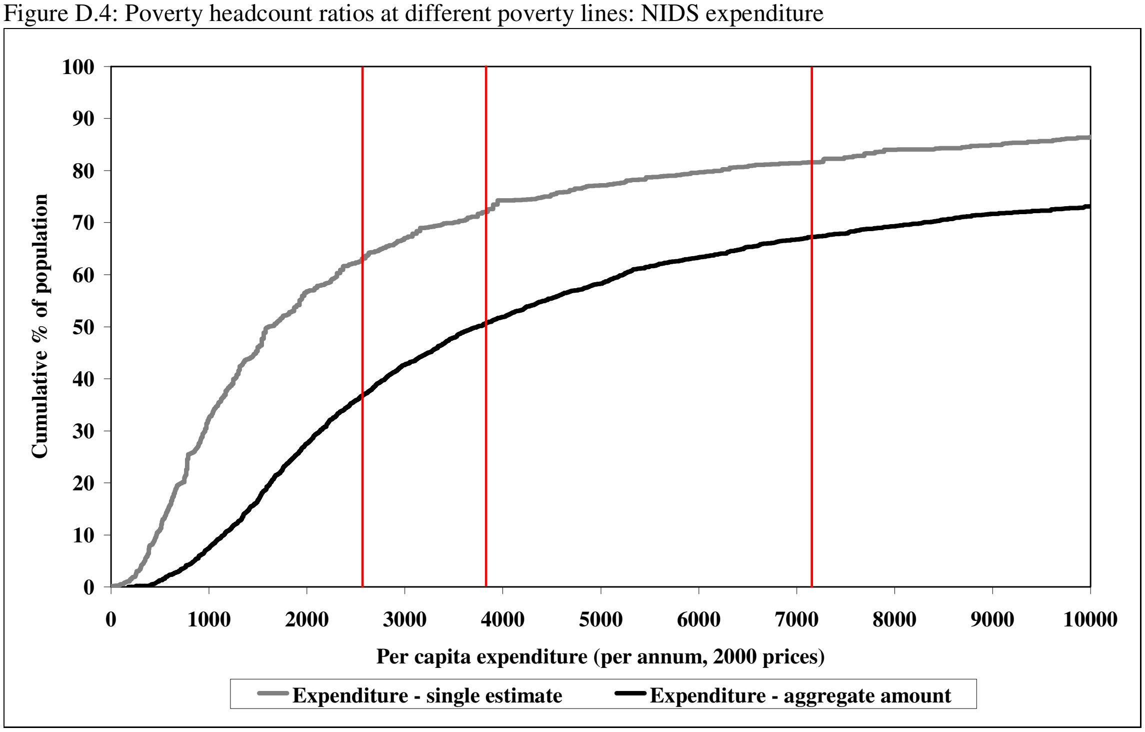


Figure D.5: Lorenz curves of per capita income (2000 prices), NIDS

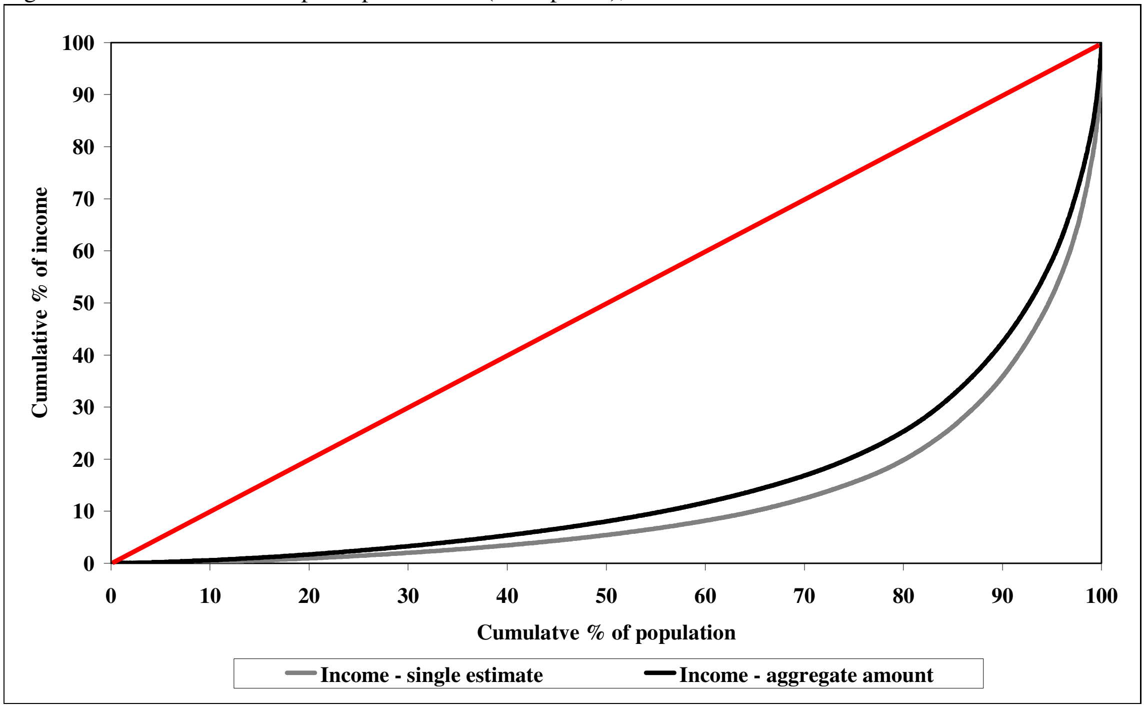
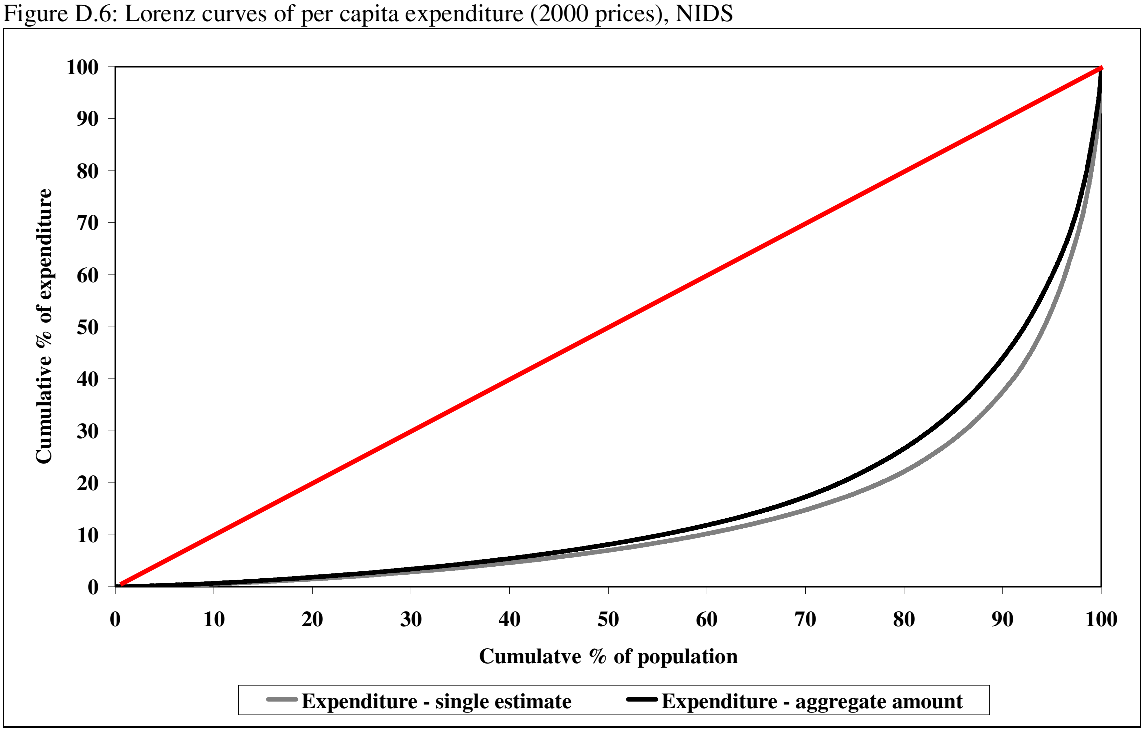


Figure D.6: Lorenz curves of per capita expenditure (2000 prices), NIDS



Appendix E: Poverty and inequality estimates in IESs and NIDS, after the application of the different intervals

Table E.1: Proportion of households in each interval, after applying different intervals to the IESs

	IES 1995	IES 2000	IES 2005/ 2006		IES 1995	IES 2000	IES 2005/ 2006
Census 1996 intervals (2000 prices)				**AMPS 2000 intervals (2000 prices)**			
None	0.0%	1.3%	0.2%	R1-R199	0.2%	4.7%	3.7%
R1 – R251	0.6%	5.6%	5.0%	R200-R299	0.8%	3.2%	3.0%
R251 – R627	11.0%	19.1%	11.6%	R300-R399	1.2%	3.7%	2.8%
R627 – R1 254	20.3%	22.8%	23.7%	R400-R499	2.4%	3.7%	3.1%
R1 254 – R1 880	14.0%	12.0%	14.5%	R500-R599	6.0%	8.3%	2.6%
R1 880 – R3 134	15.6%	13.1%	13.9%	R600-R699	3.8%	4.8%	6.3%
R3 134 – R4 387	8.9%	6.5%	6.8%	R700-R799	3.4%	4.1%	4.0%
R4 387 – R5 641	6.1%	3.7%	4.5%	R800-R899	3.2%	4.0%	3.8%
R5 641 – R7 521	6.2%	4.3%	4.5%	R900-R999	2.5%	2.9%	3.5%
R7 521 – R10 028	5.7%	3.3%	4.0%	R1 000-R1 099	2.3%	3.1%	3.2%
R10 028 – R13 788	4.8%	3.2%	4.0%	R1 100-R1 199	4.7%	3.7%	2.9%
R13 788 – R20 055	4.0%	2.6%	3.2%	R1 200-R1 399	5.4%	5.3%	5.9%
R20 055 – R37 603	2.2%	2.0%	2.8%	R1 400-R1 599	4.0%	4.2%	5.1%
R37 603+	0.8%	0.5%	1.3%	R1 600-R1 999	7.9%	6.3%	7.2%
	100.0%	100.0%	100.0%	R2 000-R2 499	7.2%	5.9%	6.2%
Census 2001 intervals (2000 prices)				R2 500-R2 999	5.3%	4.4%	4.8%
None	1.9%	1.3%	0.2%	R3 000-R3 999	7.5%	6.2%	6.1%
R1 – R377	14.4%	9.9%	8.6%	R4 000-R4 999	5.8%	3.7%	4.1%
R377 – R754	21.6%	20.3%	14.5%	R5 000-R5 999	4.3%	2.5%	3.1%
R754 – R1 509	22.3%	23.1%	24.5%	R6 000-R6 999	3.5%	2.4%	2.5%
R1 509 – R3 017	17.6%	18.5%	20.5%	R7 000-R7 999	2.6%	1.7%	2.0%
R3 017 – R6 035	13.4%	12.1%	13.2%	R8 000-R8 999	2.6%	1.6%	1.5%
R6 035 – R12 070	6.8%	8.4%	9.8%	R9 000-R9 999	1.8%	1.0%	1.5%
R12 070 – R24 140	1.5%	4.8%	5.8%	R10 000-R10 999	1.6%	1.2%	1.2%
R24 140 – R48 279	0.4%	1.4%	2.4%	R11 000-R11 999	1.5%	0.9%	1.1%
R48 279 – R96 558	0.1%	0.2%	0.5%	R12 000-R13 999	2.1%	1.5%	1.9%
R96 588 – R193 117	0.0%	0.1%	0.2%	R14 000-R15 999	1.5%	1.0%	1.5%
R193 117+	1.9%	0.0%	0.0%	R16 000-R17 999	1.3%	0.9%	0.8%
	100.0%	100.0%	100.0%	R18 000-R19 999	0.8%	0.6%	0.8%
GHS 2009 intervals (2000 prices)				R20 000+	3.0%	2.6%	4.1%
R0	0.0%	1.3%	0.2%		100.0%	100.0%	100.0%
R1 – R115	0.1%	1.8%	1.4%				
R115 – R231	0.4%	3.0%	2.9%				
R231 – R461	3.2%	8.5%	6.9%				
R461 – R692	10.6%	14.0%	9.7%				
R692 – R1 038	10.2%	12.8%	12.8%				
R1 038 – R1 442	12.3%	11.5%	12.1%				
R1 442 – R2 884	22.5%	19.1%	21.0%				
R2 884 – R5 767	17.7%	12.4%	13.7%				
R5 767+	23.1%	15.7%	19.4%				
	100.0%	100.0%	100.0%				

Table E.1: Continued

	IES 1995	IES 2000	IES 2005/ 2006		IES 1995	IES 2000	IES 2005/ 2006
R500 intervals				**R1 000 intervals**			
R0 – R499	13.4%	15.2%	8.8%	R0 – R999	34.4%	39.4%	23.5%
R500 – R999	21.1%	24.2%	14.7%	R1 000 – R1 999	22.7%	22.6%	24.4%
R1 000 – R1 499	14.0%	13.9%	13.8%	R2 000 – R2 999	11.1%	10.3%	12.7%
R1 500 – R1 999	8.7%	8.7%	10.6%	R3 000 – R3 999	7.1%	6.2%	7.7%
R2 000 – R2 499	6.2%	5.9%	7.4%	R4 000 – R4 999	5.0%	3.7%	4.7%
R2 500 – R2 999	4.9%	4.4%	5.4%	R5 000 – R5 999	3.9%	2.5%	3.5%
R3 000 – R3 499	4.0%	3.4%	4.1%	R6 000 – R6 999	3.0%	2.4%	2.7%
R3 500 – R3 999	3.1%	2.8%	3.6%	R7 000 – R7 999	2.2%	1.7%	2.2%
R4 000 – R4 499	2.8%	2.0%	2.4%	R8 000 – R8 999	1.9%	1.6%	2.0%
R4 500 – R4 999	2.2%	1.6%	2.3%	R9 000 – R9 999	1.4%	1.0%	1.6%
R5 000 – R5 499	2.1%	1.4%	2.0%	R10 000 – R10 999	1.2%	1.2%	1.4%
R5 500 – R5 999	1.8%	1.1%	1.5%	R11 000 – R11 999	1.0%	0.9%	1.1%
R6 000 – R6 499	1.7%	1.4%	1.5%	R12 000 – R12 999	0.9%	0.8%	1.1%
R6 500 – R6 999	1.3%	1.1%	1.2%	R13 000 – R13 999	0.5%	0.7%	0.9%
R7 000 – R7 499	1.1%	1.1%	1.2%	R14 000 – R14 999	0.5%	0.6%	1.0%
R7 500 – R7 999	1.1%	0.7%	1.0%	R15 000 – R15 999	0.4%	0.5%	0.7%
R8 000 – R8 499	1.1%	0.9%	1.0%	R16 000 – R16 999	0.4%	0.5%	0.7%
R8 500 – R8 999	0.9%	0.7%	1.1%	R17 000 – R17 999	0.3%	0.4%	0.7%
R9 000 – R9 499	0.8%	0.6%	0.8%	R18 000 – R18 999	0.2%	0.4%	0.7%
R9 500 – R9 999	0.6%	0.5%	0.9%	R19 000 – R19 999	0.2%	0.2%	0.6%
R10 000 – R10 499	0.6%	0.5%	0.7%	R20 000+	1.5%	2.6%	6.1%
R10 500 – R10 999	0.6%	0.6%	0.7%		100.0%	100.0%	100.0%
R11 000 – R11 499	0.5%	0.5%	0.6%	**R2 000 intervals**			
R11 500 – R11 999	0.5%	0.3%	0.4%	R0 – R1 999	57.2%	62.0%	47.9%
R12 000 – R12 499	0.5%	0.3%	0.5%	R2 000 – R3 999	18.2%	16.5%	20.4%
R12 500 – R12 999	0.5%	0.4%	0.6%	R4 000 – R5 999	9.0%	6.2%	8.2%
R13 000 – R13 499	0.3%	0.4%	0.5%	R6 000 – R7 999	5.3%	4.2%	4.9%
R13 500 – R13 999	0.2%	0.3%	0.4%	R8 000 – R9 999	3.3%	2.6%	3.6%
R14 000 – R14 499	0.3%	0.3%	0.5%	R10 000 – R11 999	2.2%	2.0%	2.5%
R14 500 – R14 999	0.2%	0.2%	0.5%	R12 000 – R13 999	1.4%	1.5%	2.0%
R15 000 – R15 499	0.2%	0.3%	0.3%	R14 000 – R15 999	0.9%	1.0%	1.8%
R15 500 – R15 999	0.2%	0.2%	0.4%	R16 000 – R17 999	0.6%	0.9%	1.4%
R16 000 – R16 499	0.2%	0.2%	0.4%	R18 000 – R19 999	0.4%	0.6%	1.3%
R16 500 – R16 999	0.2%	0.3%	0.3%	R20 000+	1.5%	2.6%	6.1%
R17 000 – R17 499	0.2%	0.3%	0.3%		100.0%	100.0%	100.0%
R17 500 – R17 999	0.1%	0.2%	0.4%				
R18 000 – R18 499	0.1%	0.1%	0.4%				
R18 500 – R18 999	0.1%	0.2%	0.3%				
R19 000 – R19 499	0.1%	0.1%	0.3%				
R19 500 – R19 999	0.1%	0.1%	0.3%				
R20 000+	1.5%	2.6%	6.1%				
	100.0%	100.0%	100.0%				

Table E.2: FGT poverty estimates, after applying different intervals on the three IESs

		FGT poverty index		
		P0	**P1**	**P2**
Poverty line: R211 per month (2000 prices)				
IES 1995				
The actual continuous income variable		0.286	0.106	0.053
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.275	0.106	0.053
	Census 1996 intervals (2000 prices)	0.271	0.109	0.056
	Census 2001 intervals (2000 prices)	0.253	0.102	0.053
	GHS 2009 intervals (2000 prices)	0.252	0.100	0.052
	R500 intervals	0.292	0.119	0.067
	R1 000 intervals	0.305	0.123	0.063
	R2 000 intervals	0.227	0.050	0.014
IES 2000				
The actual continuous income variable		0.429	0.206	0.127
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.422	0.202	0.123
	Census 1996 intervals (2000 prices)	0.417	0.207	0.129
	Census 2001 intervals (2000 prices)	0.412	0.198	0.123
	GHS 2009 intervals (2000 prices)	0.411	0.198	0.123
	R500 intervals	0.416	0.192	0.114
	R1 000 intervals	0.426	0.199	0.116
	R2 000 intervals	0.391	0.127	0.055
IES 2005/2006				
The actual continuous income variable		0.338	0.137	0.075
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.326	0.137	0.075
	Census 1996 intervals (2000 prices)	0.324	0.133	0.073
	Census 2001 intervals (2000 prices)	0.317	0.132	0.073
	GHS 2009 intervals (2000 prices)	0.319	0.133	0.074
	R500 intervals	0.341	0.138	0.076
	R1 000 intervals	0.332	0.140	0.077
	R2 000 intervals	0.354	0.128	0.060
Poverty line: R322 per month (2000 prices)				
IES 1995				
The actual continuous income variable		0.434	0.195	0.111
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.433	0.195	0.111
	Census 1996 intervals (2000 prices)	0.446	0.194	0.112
	Census 2001 intervals (2000 prices)	0.406	0.187	0.107
	GHS 2009 intervals (2000 prices)	0.430	0.187	0.106
	R500 intervals	0.419	0.203	0.123
	R1 000 intervals	0.408	0.208	0.125
	R2 000 intervals	0.398	0.149	0.065
IES 2000				
The actual continuous income variable		0.559	0.307	0.204
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.562	0.303	0.200
	Census 1996 intervals (2000 prices)	0.569	0.306	0.205
	Census 2001 intervals (2000 prices)	0.538	0.297	0.197
	GHS 2009 intervals (2000 prices)	0.551	0.295	0.197
	R500 intervals	0.559	0.296	0.192
	R1 000 intervals	0.553	0.300	0.195
	R2 000 intervals	0.497	0.241	0.133

Table E.2: Continued

		FGT poverty index		
		P0	**P1**	**P2**
IES 2005/2006				
The actual continuous income variable		0.488	0.234	0.141
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.484	0.234	0.141
	Census 1996 intervals (2000 prices)	0.501	0.227	0.137
	Census 2001 intervals (2000 prices)	0.464	0.225	0.135
	GHS 2009 intervals (2000 prices)	0.482	0.225	0.136
	R500 intervals	0.488	0.235	0.142
	R1 000 intervals	0.484	0.235	0.143
	R2 000 intervals	0.472	0.228	0.130
Poverty line: R593 per month (2000 prices)				
IES 1995				
The actual continuous income variable		0.622	0.352	0.236
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.625	0.351	0.235
	Census 1996 intervals (2000 prices)	0.612	0.350	0.234
	Census 2001 intervals (2000 prices)	0.623	0.339	0.226
	GHS 2009 intervals (2000 prices)	0.605	0.340	0.226
	R500 intervals	0.618	0.356	0.243
	R1 000 intervals	0.612	0.358	0.245
	R2 000 intervals	0.606	0.322	0.197
IES 2000				
The actual continuous income variable		0.710	0.462	0.342
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.713	0.458	0.339
	Census 1996 intervals (2000 prices)	0.705	0.459	0.340
	Census 2001 intervals (2000 prices)	0.717	0.448	0.331
	GHS 2009 intervals (2000 prices)	0.695	0.448	0.331
	R500 intervals	0.701	0.454	0.333
	R1 000 intervals	0.709	0.455	0.334
	R2 000 intervals	0.706	0.417	0.284
IES 2005/2006				
The actual continuous income variable		0.657	0.395	0.275
Applying the intervals on the income data	AMPS 2000 intervals (2000 prices)	0.659	0.394	0.274
	Census 1996 intervals (2000 prices)	0.651	0.389	0.268
	Census 2001 intervals (2000 prices)	0.665	0.381	0.264
	GHS 2009 intervals (2000 prices)	0.639	0.381	0.264
	R500 intervals	0.663	0.396	0.275
	R1 000 intervals	0.665	0.395	0.275
	R2 000 intervals	0.656	0.388	0.266

Table E.3: FGT poverty estimates and Gini coefficients, after applying different intervals on NIDS

		Poverty line: R211 per month (2000 prices)			Poverty line: R2322 per month (2000 prices)			Poverty line: R593 per month (2000 prices)			Gini
		P0	**P1**	**P2**	**P0**	**P1**	**P2**	**P0**	**P1**	**P2**	
Single-estimate income variable (n = 5920)											
The actual continuous income variable		0.493	0.259	0.171	0.614	0.363	0.255	0.728	0.505	0.392	0.753
Applying intervals on the data	AMPS 2007	0.484	0.247	0.156	0.600	0.350	0.241	0.728	0.493	0.379	0.721
	CS 2007	0.481	0.255	0.172	0.597	0.353	0.250	0.713	0.496	0.383	0.752
	GHS 2007	0.477	0.236	0.146	0.582	0.339	0.231	0.701	0.484	0.369	0.774
Aggregated income variable (n = 5920)											
The actual continuous income variable		0.300	0.120	0.065	0.468	0.212	0.125	0.651	0.377	0.255	0.693
Applying intervals on the data	AMPS 2007	0.299	0.118	0.064	0.461	0.210	0.123	0.657	0.373	0.253	0.687
	CS 2007	0.285	0.117	0.066	0.444	0.204	0.122	0.625	0.362	0.245	0.694
	GHS 2007	0.289	0.115	0.064	0.433	0.202	0.120	0.618	0.360	0.244	0.737
Single-estimate expenditure variable (n = 5799)											
The actual continuous expenditure variable		0.625	0.355	0.238	0.726	0.468	0.342	0.815	0.611	0.493	0.725
Applying intervals on the data	AMPS 2007	0.604	0.340	0.225	0.710	0.453	0.327	0.811	0.597	0.479	0.717
	CS 2007	0.635	0.367	0.254	0.719	0.473	0.353	0.813	0.613	0.498	0.748
	GHS 2007	0.612	0.335	0.221	0.694	0.446	0.323	0.798	0.590	0.472	0.845
Aggregated expenditure variable (n = 5799)											
The actual continuous expenditure variable		0.363	0.144	0.075	0.509	0.247	0.147	0.672	0.410	0.287	0.684
Applying intervals on the data	AMPS 2007	0.361	0.143	0.075	0.506	0.246	0.146	0.677	0.408	0.285	0.675
	CS 2007	0.346	0.137	0.073	0.488	0.234	0.140	0.649	0.394	0.273	0.691
	GHS 2007	0.361	0.143	0.076	0.483	0.241	0.146	0.648	0.400	0.280	0.702

Figure E.1: Poverty headcount ratios at different poverty lines in IES 2000 income (STC approach), after applying the Census and GHS intervals

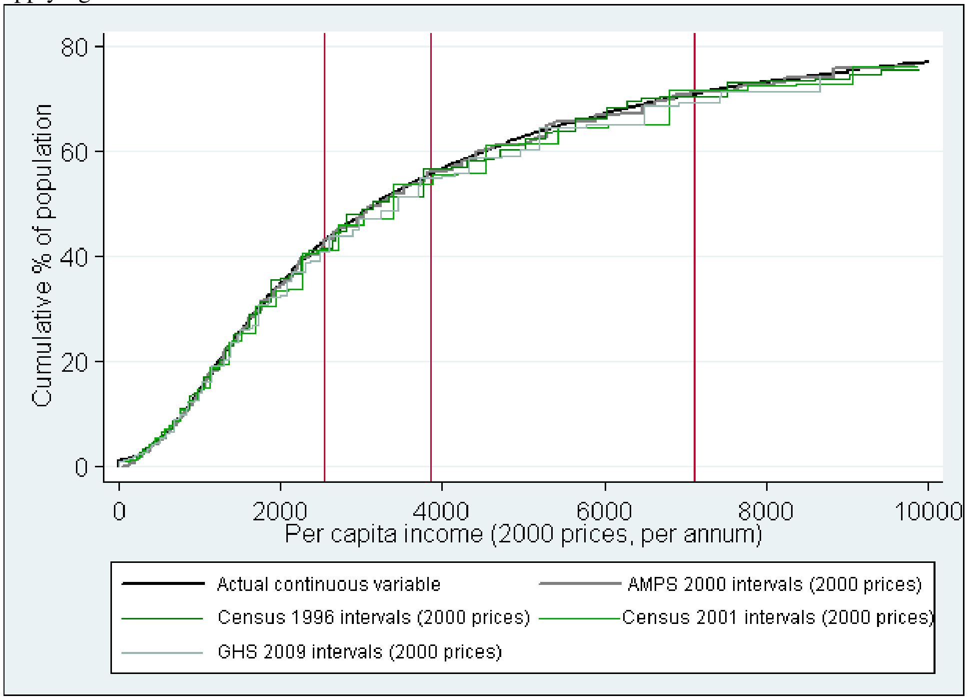


Figure E.2: Poverty headcount ratios at different poverty lines in IES 2000 income (STC approach), after applying various intervals

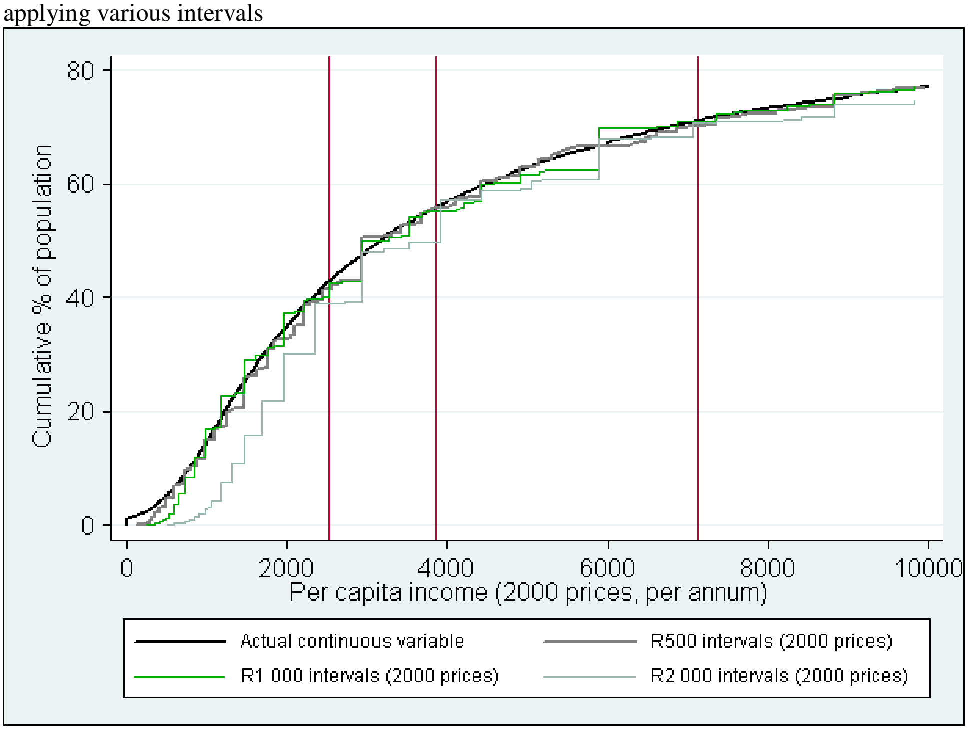


Figure E.3: Poverty headcount ratios, after applying various intervals on NIDS single estimate income and expenditure variables (Poverty line: R322 per month 2000 prices)

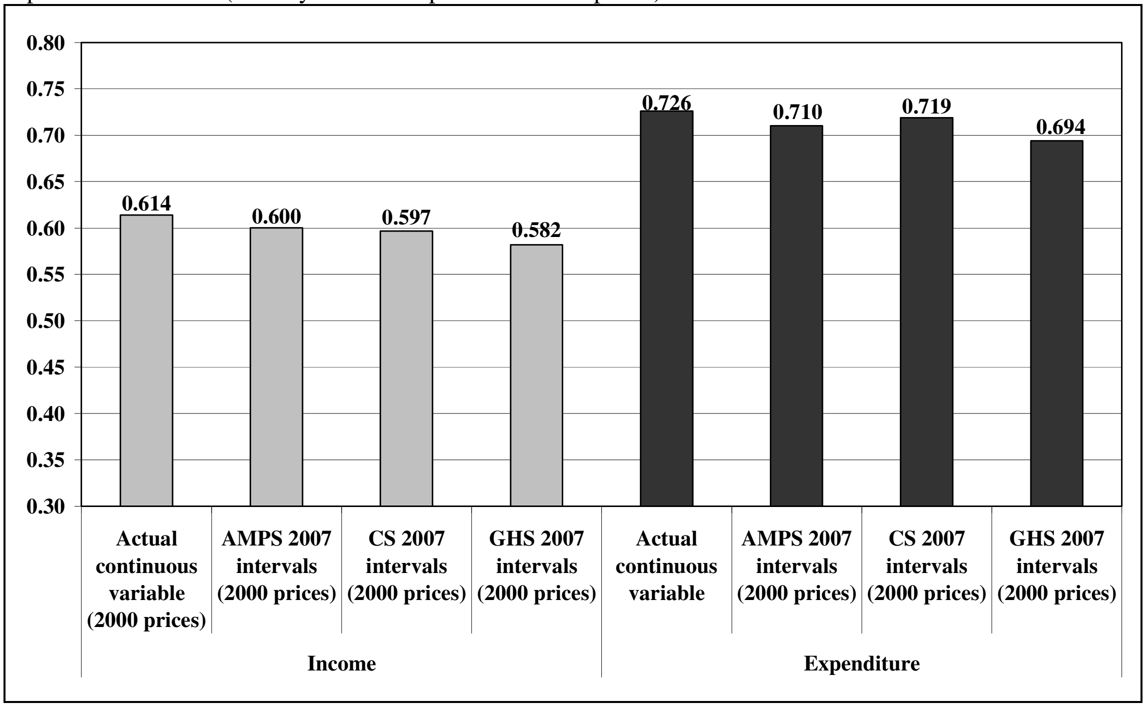


Figure E.4: Gini coefficients, after applying various intervals on NIDS single estimate income and expenditure variables (Poverty line: R322 per month 2000 prices)

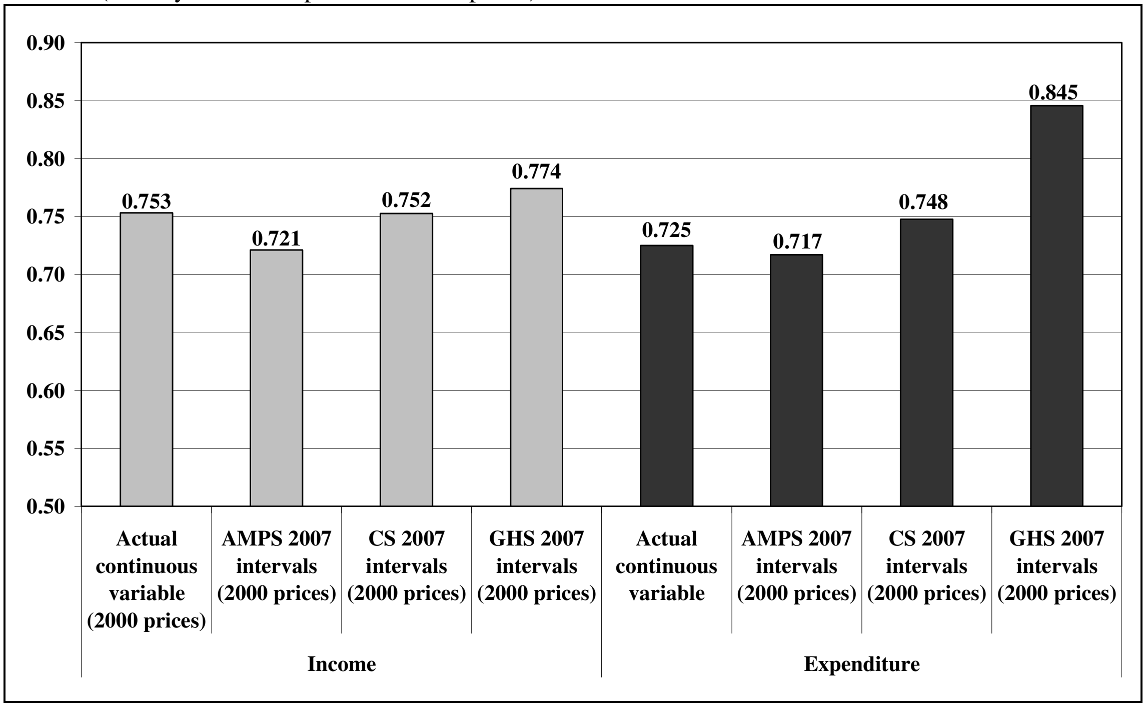


Figure E.5: Poverty headcount ratios, after applying various intervals on NIDS aggregated income and expenditure variables

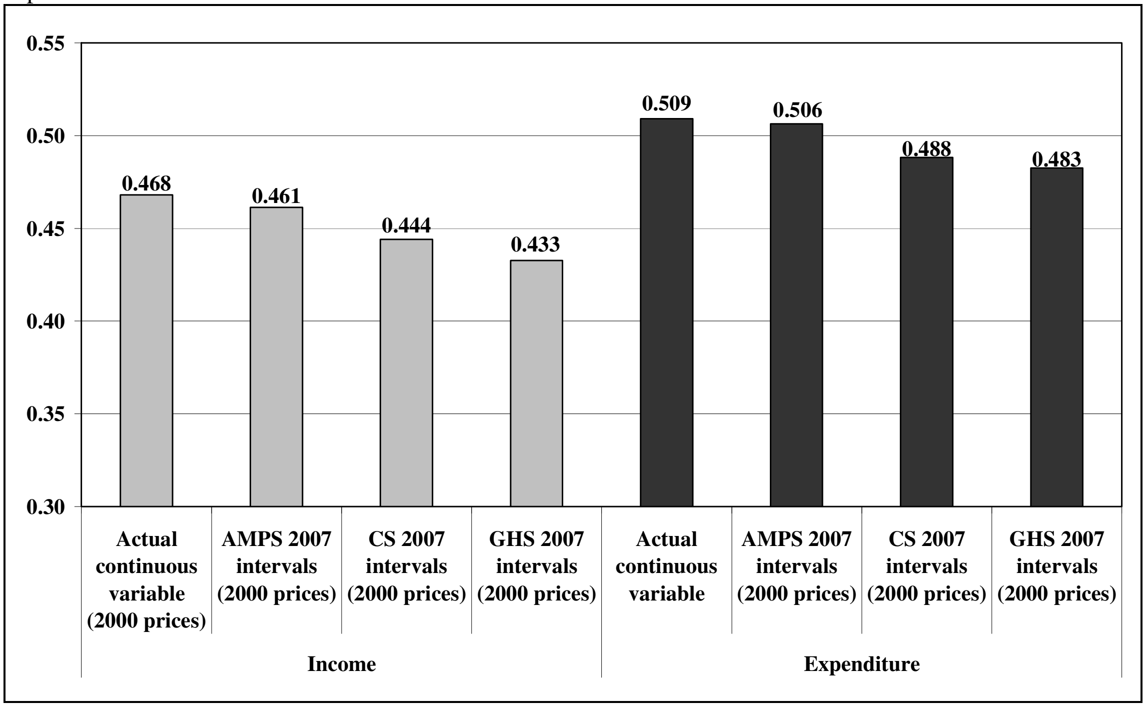
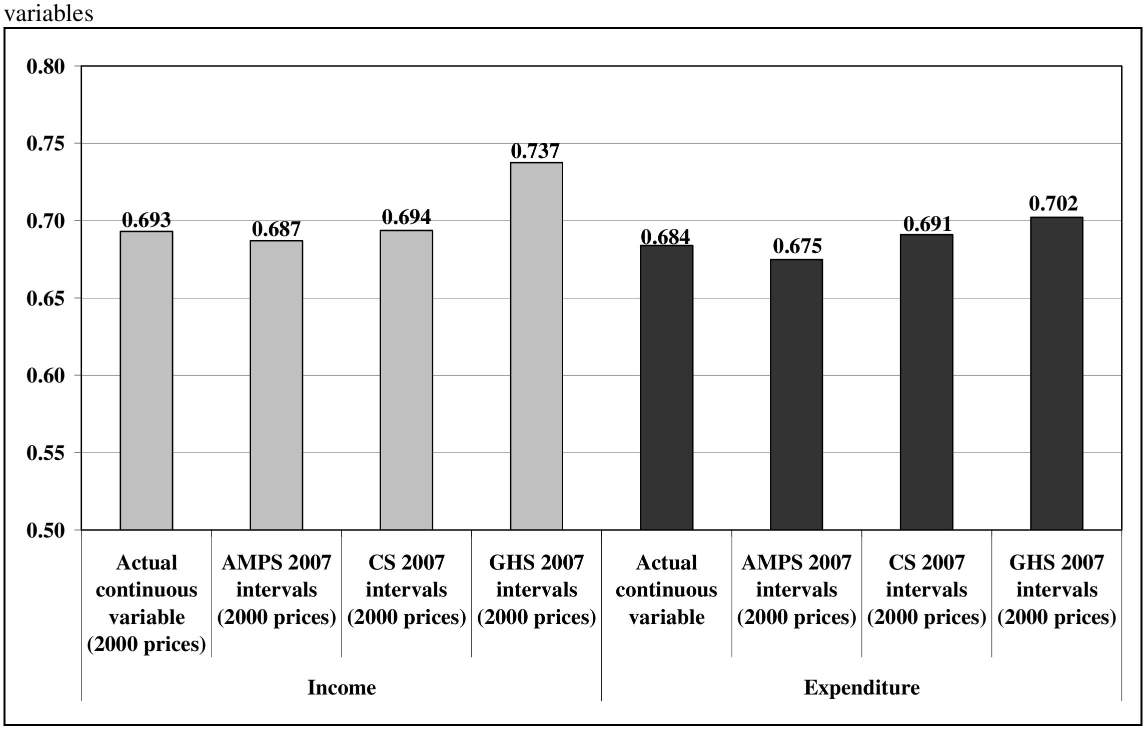


Figure E.6: Gini coefficients, after applying various intervals on NIDS aggregated income and expenditure variables



Appendix F: Poverty and inequality estimates in each survey, after adjusting the survey means in line with national accounts mean

Figure F.1: Poverty headcount before and after adjustments in line with national accounts, using per capita variables of the AMPSs (Poverty line: R322 per month 2000 prices)

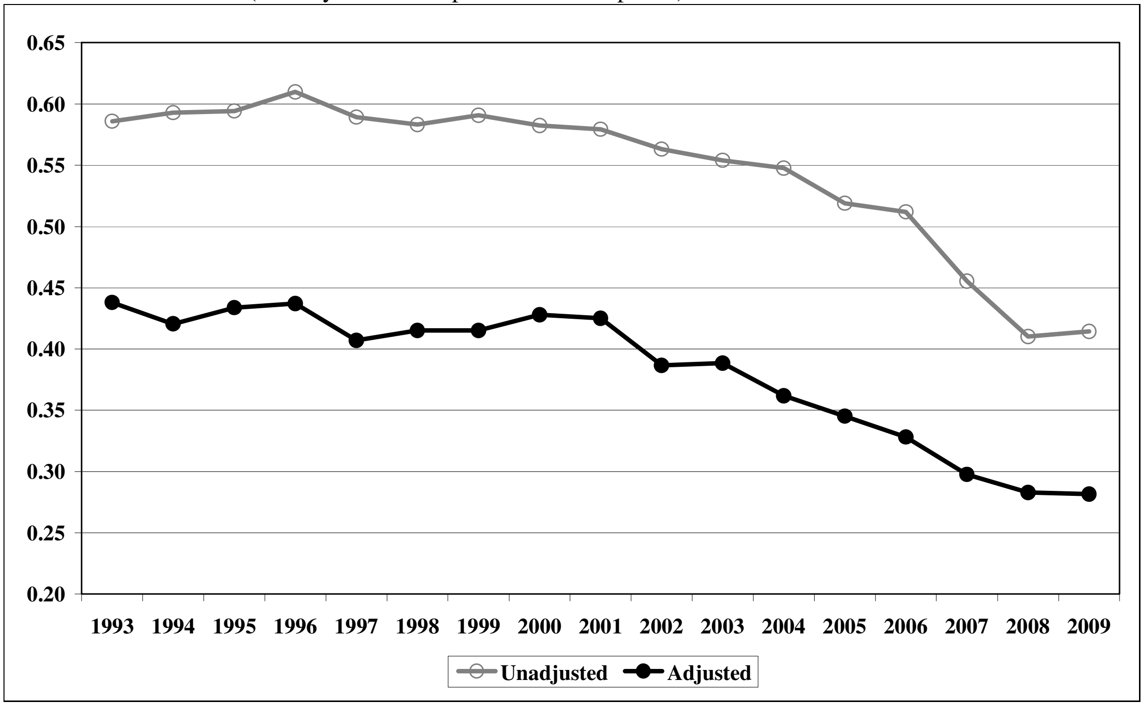


Figure F.2: Poverty headcount before and after adjustments in line with national accounts, using per capita variables of the censuses and CS 2007 (Poverty line: R322 per month 2000 prices)

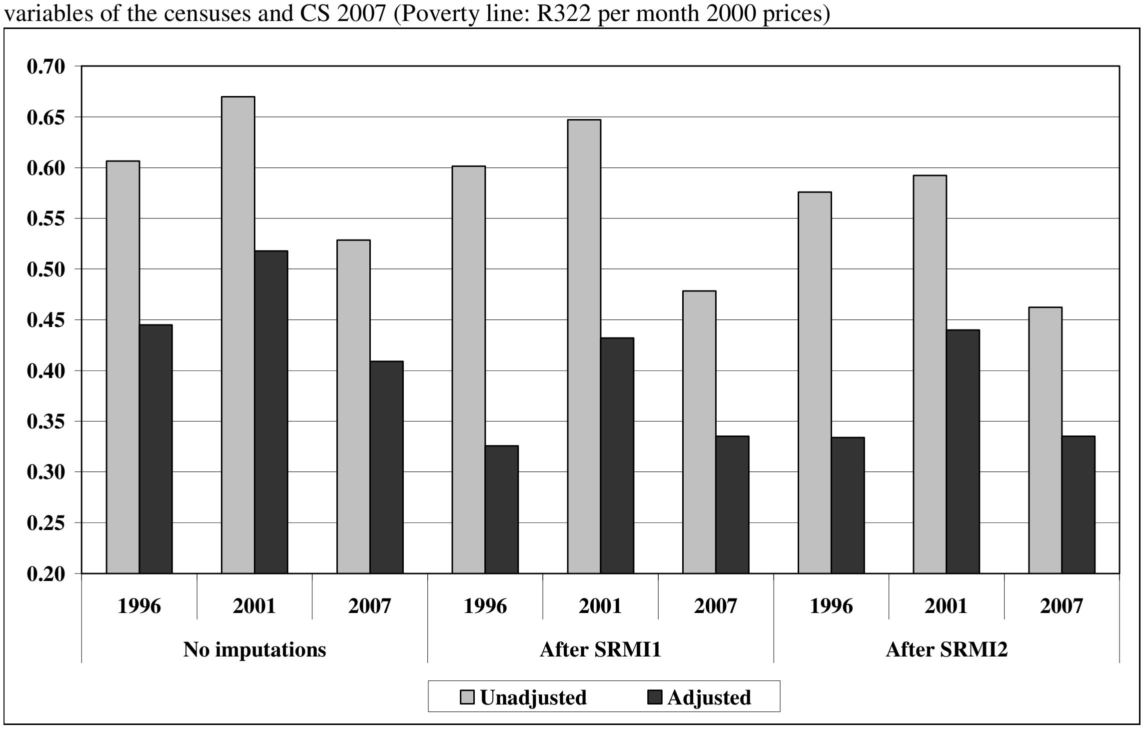


Figure F.3: Poverty headcount before and after adjustments in line with national accounts, using per capita variables of the IESs (Poverty line: R322 per month 2000 prices)

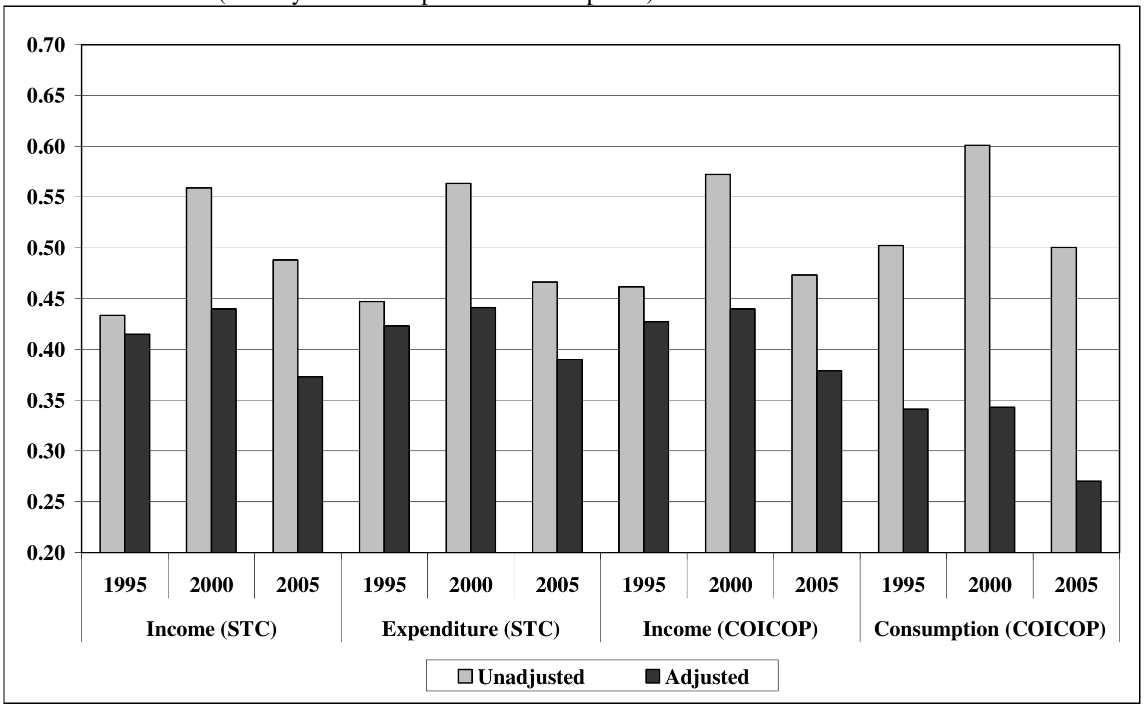


Figure F.4: Poverty headcount before and after adjustments in line with national accounts, using per capita expenditure variables of the OHSs and LFSs (Poverty line: R322 per month 2000 prices)

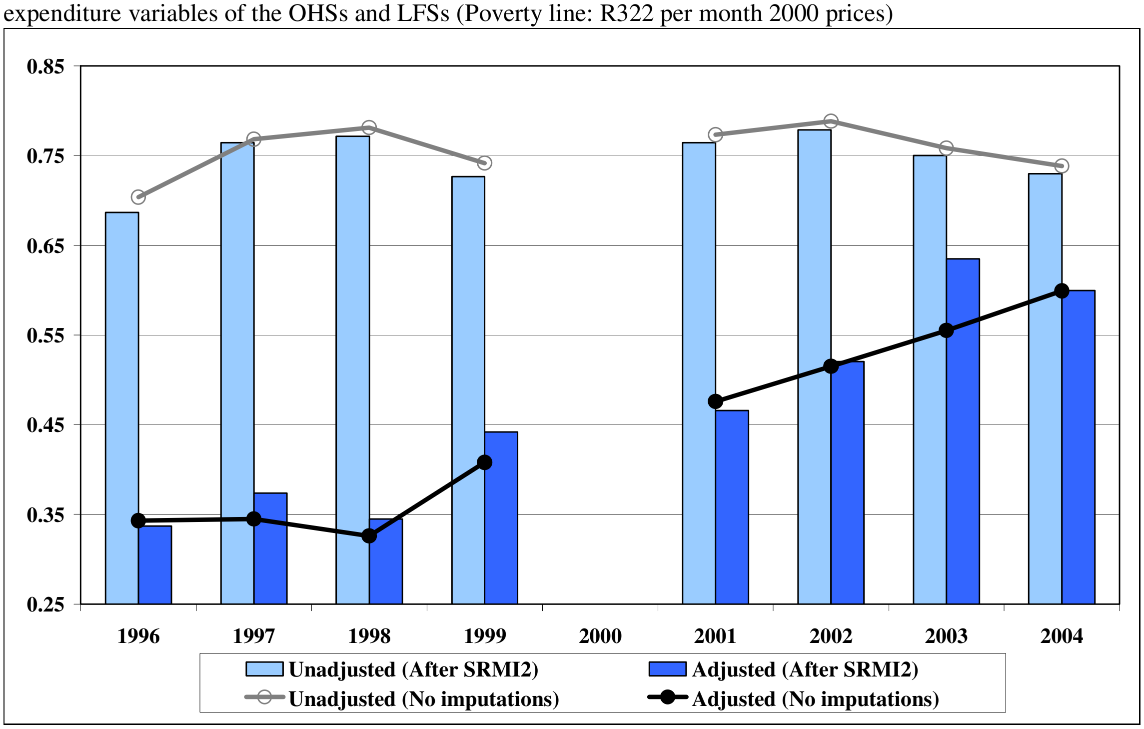


Figure F.5: Poverty headcount before and after adjustments in line with national accounts, using per capita expenditure variables of the GHSs (Poverty line: R322 per month 2000 prices

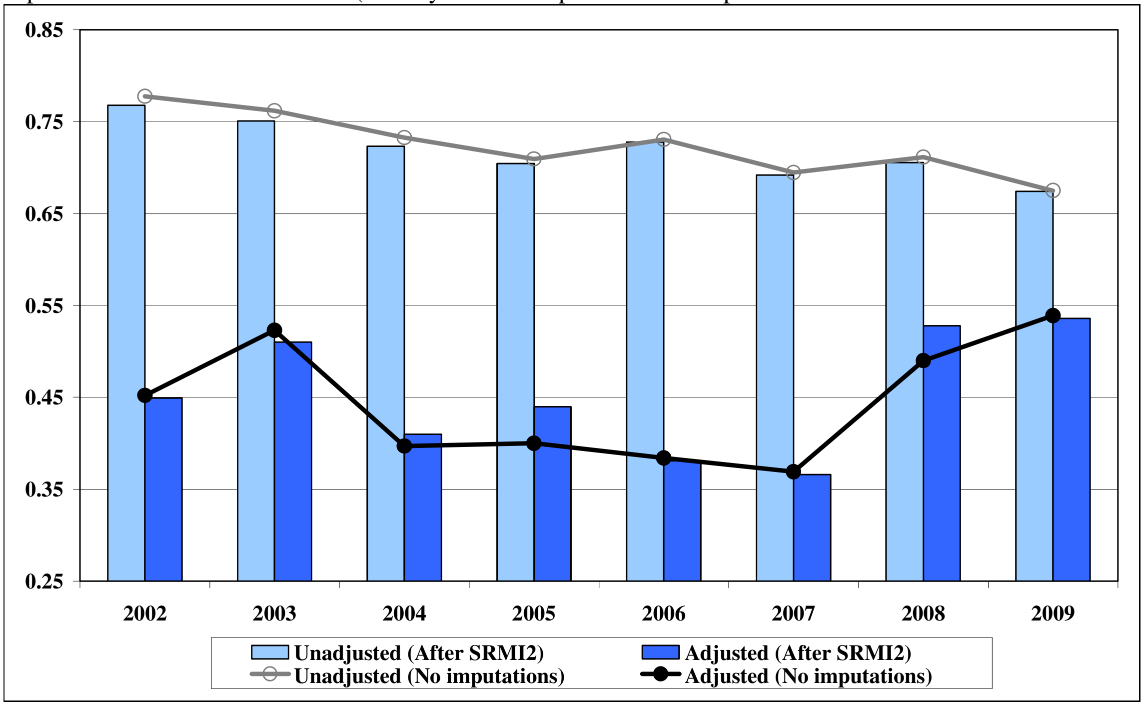
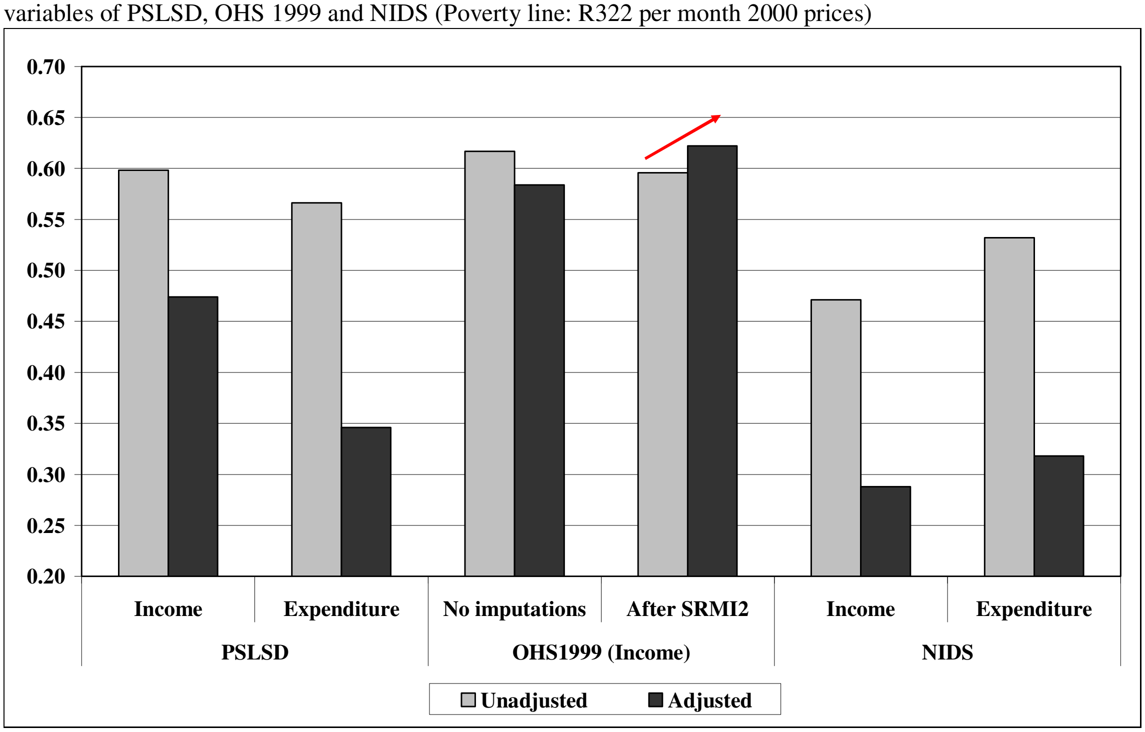


Figure F.6: Poverty headcount before and after adjustments in line with national accounts, using per capita variables of PSLSD, OHS 1999 and NIDS (Poverty line: R322 per month 2000 prices)



Appendix G: Poverty and inequality estimates in each survey, after the cross entropy approach was applied

Figure G.1: Poverty headcount before and after cross entropy approach was applied, using per capita variables of the censuses and CS 2007 (Poverty line: R322 per month 2000 prices)

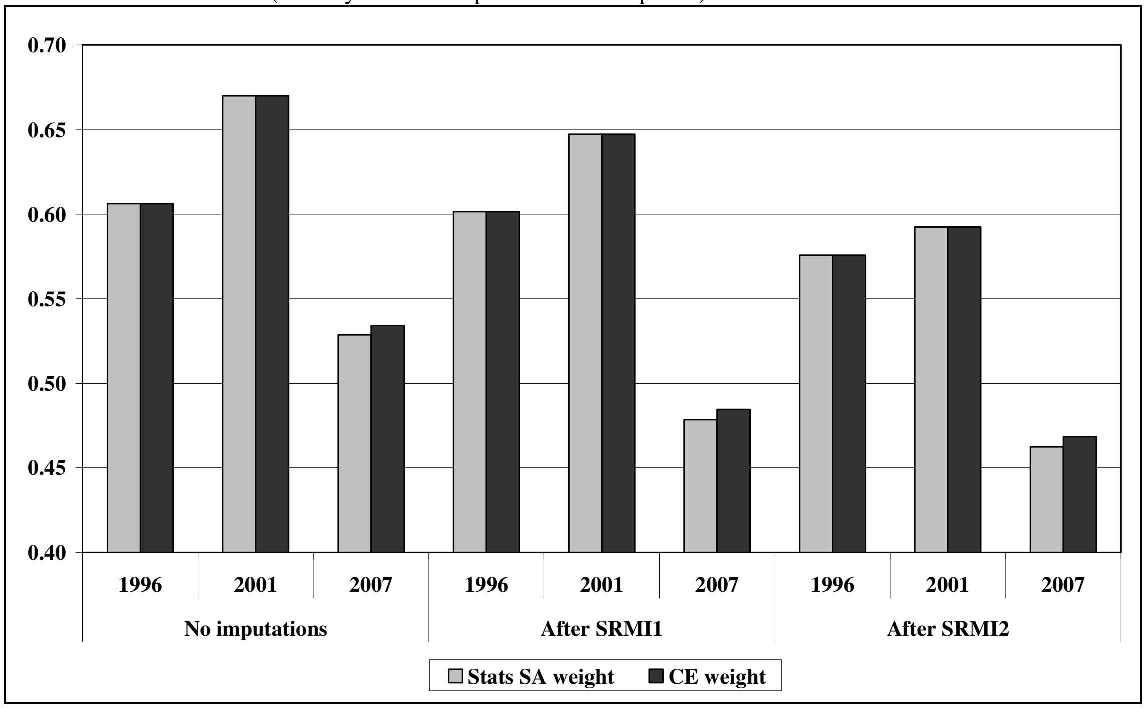


Figure G.2: Poverty headcount before and after cross entropy approach was applied, using per capita variables of the IESs (Poverty line: R322 per month 2000 prices)

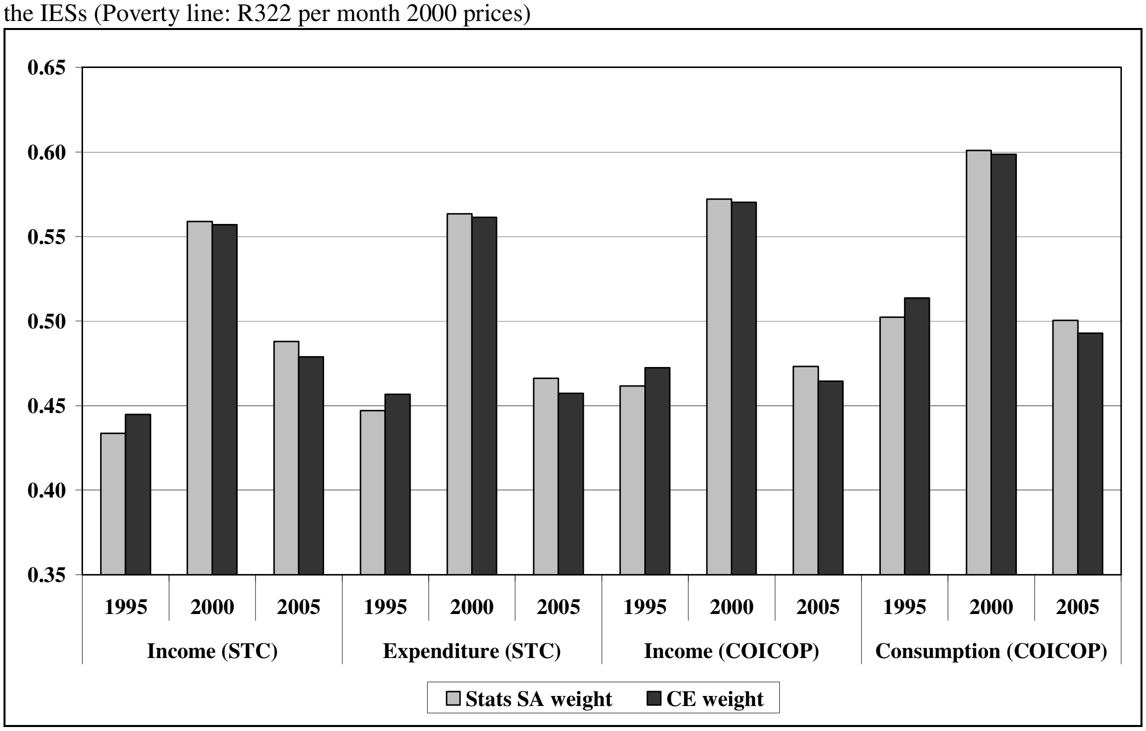


Figure G.3: Poverty headcount before and after cross entropy approach was applied, using per capita expenditure variables of the OHSs and LFSs (Poverty line: R322 per month 2000 prices)

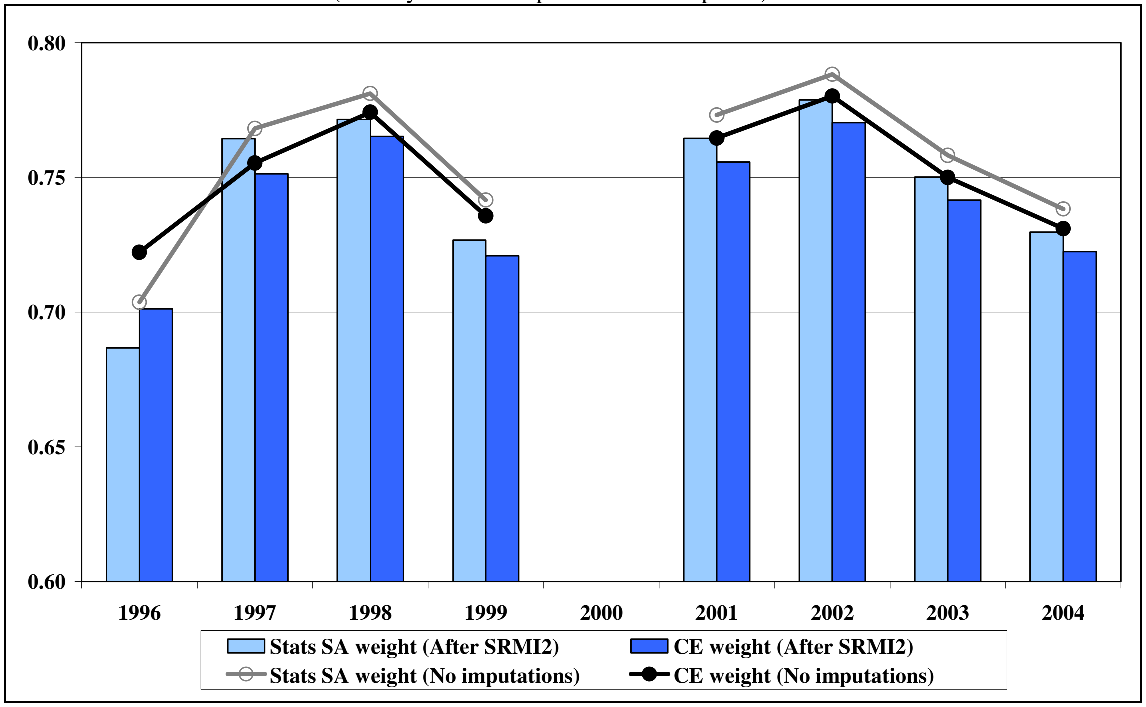


Figure G.4: Poverty headcount before and after cross entropy approach was applied, using per capita expenditure variables of the GHSs (Poverty line: R322 per month 2000 prices)