The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys

DEREK YU

Stellenbosch Economic Working Papers: 17/07

KEYWORDS: SOUTH AFRICA, HOUSEHOLD SURVEY; LABOUR MARKET TRENDS; EARNINGS
JEL: J00

DEREK YU
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: DEREKY@SUN.AC.ZA

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

BER
BUREAU FOR ECONOMIC RESEARCH

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys[1]

DEREK YU

ABSTRACT

Statistics South Africa (Stats SA) has been collecting labour market data with household surveys and in a fairly comparable format since 1993. These datasets have been studied and compared extensively in order to better understand the workings of the South African labour market. Many of these studies compare household surveys of different periods in order to identify trends, but the validity of such trends is conditional on the comparability of the different datasets. Besides, the naïve comparisons of the different datasets have been questioned. Other problems include inconsistencies in questionnaire design, coding errors, changes in the sampling frame, the oversampling of agricultural workers in OHS1995, the oversampling of subsistence agricultural workers in LFS2000a and LFS2000b, as well as the oversampling of informal workers in LFS2001a.

Most of these issues have received attention in papers by Burger and Yu (2006), Casale, Muller and Posel (2005), and Wittenberg (2004). By drawing attention to a few of the lesser known problems, this paper aims to build on the existing literature by further stimulating debate around the strengths and weaknesses of the existing survey data, as well as considering the best ways in which to analyse the existing data. The inconsistencies that occur in the data independently of the way in which questions are asked by the interview, as well as the inconsistencies that result from the way in which the survey questions are formulated or placed in a given sequence are discussed. Where possible, adjustments that may contribute towards increased consistency in the responses are suggested. Ultimately, it is hoped that the lessons learnt from such discussions will serve to inform questionnaire design in future.

Keywords: South Africa, Household Survey; Labour Market Trends; Earnings
JEL codes: J00

[1] The author gratefully acknowledges the research assistance of Paula Armstrong, and valuable comments by Rulof Burger, Hassan Essop and Servaas van der Berg.

The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys

1. INTRODUCTION

Statistics South Africa (Stats SA) has been collecting labour market data in a fairly comparable format since 1993 with the October Household Survey (OHS), which was conducted annually between 1993 and 1999[2], and the Labour Force Survey (LFS)[3], which is a biannual survey introduced in 2000 to replace the OHS[4]. These datasets have been studied and compared extensively in order to better understand the workings of the South African labour market. Many of these studies compare household surveys of different periods in order to identify trends, but the validity of such trends is conditional on the comparability of the different datasets. Besides, the naïve comparisons of the different datasets have been questioned. Other problems include inconsistencies in the questionnaire design, coding errors, changes in the sampling frame, the oversampling of agricultural workers in OHS1995, the oversampling of subsistence agricultural workers in LFS2000a and LFS2000b, as well as the oversampling of informal workers in LFS2001a.

Most of these issues have received attention in papers by Burger and Yu (2006), Casale, Muller and Posel (2005), and Wittenberg (2004). By drawing attention to a few of the lesser known problems, this paper aims to build on the existing literature by further stimulating debate around the strengths and weaknesses of the existing survey data, as well as considering the best ways in which to analyse the existing data. Where possible, ways in which the worst problems can be avoided will be suggested. Ultimately, it is hoped that the lessons learnt from such discussions will serve to inform questionnaire design in future.

In this paper, the data sources used in the analysis are the 1993 – 1999 OHS and the 2000 – 2006 LFS. Data from OHS1993 to OHS1994 are weighted using the 1991 Census weights, while data from OHS1995 to LFS2000a are weighted using the 1996 Census weights. Finally, the data from LFS2000b to LFS2006b are weighted using the 2001 Census weights.

Section 2 looks at the unweighted sample size and the weighted population size. Section 3 focuses on the inconsistencies that occur in the data independently of the way in which questions are asked by the interviewer, discussing the inconsistencies in the demographic, educational attainment, employment and earnings variables. Section 4 discusses inconsistencies in the data that result from the way in which the survey questions are formulated or placed in a given sequence. Section 5 suggests adjustments that may contribute towards increased consistency in the responses.

[2] The 1996 OHS was actually conducted in November "since enumeration for the 1996 population census took place during that time" (Stats SA 1999).
[3] The first round of LFS takes place in March and the second round in September. The only exception is the first round of the 2000 LFS, which actually took place in February.
[4] For the remainder of the paper, the OHSs will be referred to as OHS1993, OHS1994, etc., while the LFSs will be referred to as LFS2000a (for the first round of LFS in 2000), LFS2000b (second round in 2000), LFS2001a, LFS2001b, and so forth.

2. SAMPLE SIZE

Table 1 below reports the number of sampled households, individuals of all ages and the working age population (i.e., population that fall between the ages of 15 and 65) in the surveys under investigation. With the exception of OHS1996 (which coincided with the 1996 Census), OHS1998 (for which funding restrictions were more severe) and LFS2000a (which was considered to be a pilot study for the newly introduced LFSs), all surveys consisted of samples of around 30,000 households. This paper focuses mainly on the working age population.

Table 1: Sample size in each survey, 1993 – 2006

Survey	Number of households	Sample size – All ages	Sample size – 15-65 years
OHS1993	30,233	136,466[5]	86,107
OHS1994	30,279	132,469	82,446
OHS1995	29,700	130,787	81,108
OHS1996	15,920	72,889	44,001
OHS1997	29,811	140,015	82,613[6]
OHS1998	18,968	82,213[7]	49,560[7]
OHS1999	26,134	106,650	65,995
LFS2000a	9,705	38,529	23,713[8]
LFS2000b	26,648	105,370	65,612[9]
LFS2001a	28,170	107,726	67,903
LFS2001b	27,356	106,439	66,517[10]
LFS2002a	29,010	109,408	69,150
LFS2002b	26,474	102,480	64,372
LFS2003a	26,702	100,834	63,825
LFS2003b	26,825	98,748	62,869
LFS2004a	26,829	98,256	62,696
LFS2004b	28,594	109,888	68,433[11]
LFS2005a	28,841	110,671	69,101
LFS2005b	28,418	109,079	68,269
LFS2006a	28,649	108,345	68,386
LFS2006b	28,363	106,900	66,867

[5] OHS1993: although the sample size (of all ages) is actually 136,468 from the person file, 2 people (aged 20 and 25 – both of them are employed) do not have a household number (i.e., unique household number is a missing value), and they are excluded from all the OHS1993 analyses for the remainder of the paper. Note that in OHS1993 the person file contains data on both the person questions and work-related questions, but from OHS1994 onwards there are two separate files, namely the person file which contains data on person questions and the worker file which contains data on work activity questions. Only people aged 15 or above at the time of the survey are allowed to answer the work activity questions

[6] OHS1997: in the person file, 82,613 people are aged between 15 and 65 years, but 6 of them are not contained in the worker data file.

[7] OHS1998: although the person data shows that there are 82,263 observations, 50 of these occur twice in the dataset (of these 50 people, 37 of them fall 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 are 49,599 people aged between 15 and 65 years. It is found that 39 people (49,599 – 49,560) only exist in the worker file but not in the person file. The reason for this is not known. All the 39 people who went missing in the person file come from Western Cape. 36 of them are Coloureds and the remaining 3 are Blacks. Finally, 21 of them (or 10,824, if weighted) are employed. It was found best to exclude these 39 people from all the OHS1998 analysis for the remainder of the paper.

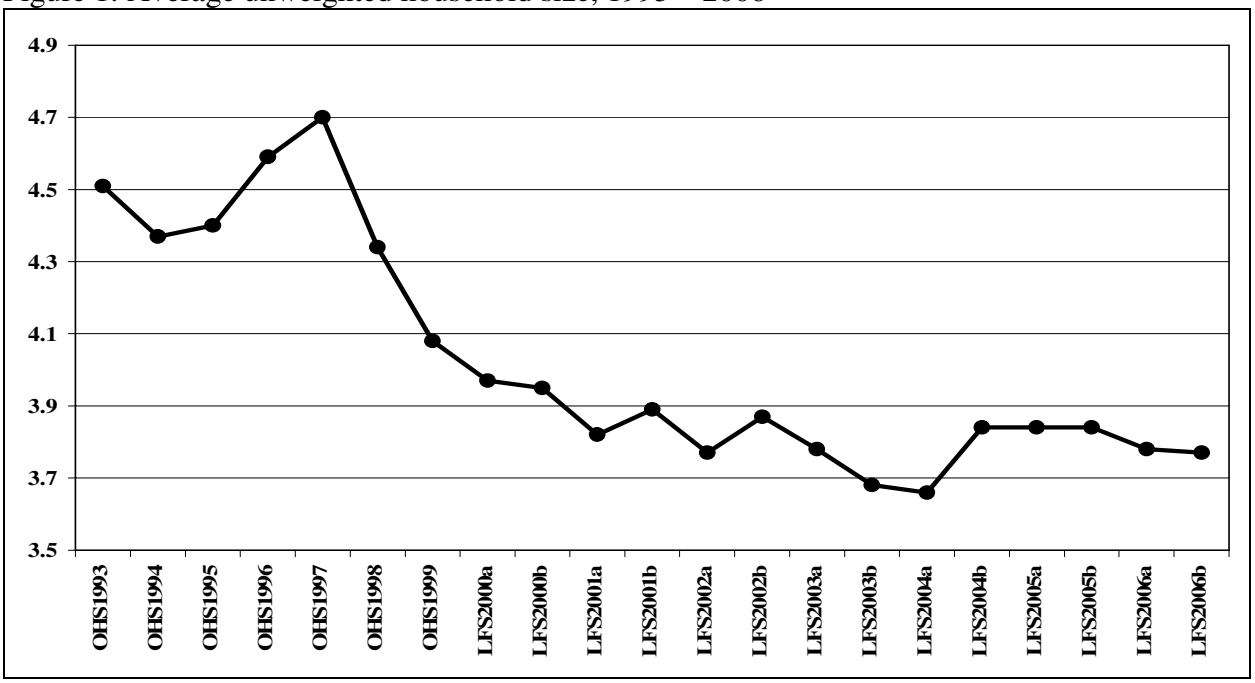
[8] LFS2000a: in the person file, 23,713 people are aged between 15 and 65 years, but 13 of them are not found in the worker file.

[9] LFS2000b: in the person file, 65,612 people are aged between 15 and 65 years, but 113 of them are not found in the worker file.

[10] LFS2001b: in the person file, 66,517 people are aged between 15 and 65 years, but 1 of them are not found in the worker file.

[11] LFS2004b: in the person file, 68,433 people are aged between 15 and 65 years, but 1 of them are not found in the worker file.

Figure 1: Average unweighted household size, 1993 – 2006



The number of individuals yielded by sampling about 30,000 households tends to decrease slightly over time, and this is caused by the downward trending (unweighted) mean household size, as shown in Figure 1 above. Table 2 shows more detail by comparing the average unweighted household size with the weighted household size.

Table 2: Average household size, unweighted vs. weighted, 1993 – 2006

	Unweighted	**Weighted**
OHS1993	4.51	N/A
OHS1994	4.37	N/A
OHS1995	4.40	N/A
OHS1996	4.59	N/A
OHS1997	4.70	4.56
OHS1998	4.34	4.28
OHS1999	4.08	3.72
LFS2000a	3.97	N/A
LFS2000b	3.95	3.94
LFS2001a	3.82	3.90
LFS2001b	3.89	3.94
LFS2002a	3.77	N/A
LFS2002b	3.87	3.90
LFS2003a	3.78	N/A
LFS2003b	3.68	3.74
LFS2004a	3.66	3.76
LFS2004b	3.84	3.69
LFS2005a	3.84	3.70
LFS2005b	3.84	N/A
LFS2006a	3.78	N/A
LFS2006b	3.77	N/A

Note: There is no household weight in OHS1993 – OHS1996, LFS2000a, LFS2002a, LFS2003a, and LFS2005b – LFS2006b.

The population size and its change between successive years provide another check on the comparability of different surveys. Table 3 indicates that the total and working age populations followed a similar trend over time, to the extent that the working age population represents a fairly stable proportion – around 62% – of the total population.

Looking at the total population, except for the obvious jump between 1993 and 1994 (as the sample was extended to include the TBVC states) and the slight decline of population growth between 1994 and 1995, a relatively larger growth of population of approximately 2% during the OHS years is observed. The growth slowed down during the changeover from the OHS to the LFS, followed by a rapid increase of 2.75% between LFS2000a and LFS2000b. From then onwards, the growth rate between successive surveys appears to have settled down at approximately 0.5%, with the exception of the zero growth rate in LFS2001b. Table A.1 in the appendix provides more detail on the total population by showing the percentage of people in each age group.

Table 3: Size and growth of total and working age populations, 1993 – 2006

	Total population		**Working age population**	
	Size	Growth rate between two successive surveys (%)	Size	Growth rate between two successive surveys (%)
OHS1993	32,207,758		19,627,903	
OHS1994	40,251,142	24.97%	24,074,568	22.65%
OHS1995	39,659,831	-1.47%	24,190,583	0.48%
OHS1996	40,582,538	2.33%	24,909,065	2.97%
OHS1997	41,443,101	2.12%	25,506,089	2.40%
OHS1998	42,211,816	1.85%	25,665,233	0.62%
OHS1999	43,271,686	2.51%	26,246,545	2.26%
LFS2000a	43,620,361	0.81%	26,465,110	0.83%
LFS2000b	44,821,345	2.75%	27,836,456	5.18%
LFS2001a	45,080,410	0.58%	28,062,004	0.81%
LFS2001b	45,081,045	0.00%	28,084,327	0.08%
LFS2002a	45,324,735	0.54%	28,298,255	0.76%
LFS2002b	45,560,990	0.52%	28,495,088	0.70%
LFS2003a	45,810,074	0.55%	28,724,521	0.81%
LFS2003b	46,046,026	0.52%	28,906,230	0.63%
LFS2004a	46,270,894	0.49%	29,099,787	0.67%
LFS2004b	46,490,122	0.47%	29,270,821	0.59%
LFS2005a	46,699,967	0.45%	29,489,763	0.75%
LFS2005b	46,917,195	0.47%	29,663,379	0.59%
LFS2006a	47,184,311	0.57%	29,817,824	0.52%
LFS2006b	47,429,106	0.52%	29,972,521	0.52%

3. COMPARISON OF VARIABLES ACROSS SURVEYS

This section discusses the inconsistencies in the data that occur for reasons other than the way in which questions are asked, or confusion that arises from the way that a question is formulated or where the question is placed within the survey questionnaire as a whole. It covers the demographic, educational attainment, employment and earnings variables.

3.1 DEMOGRAPHIC VARIABLES

3.1.1 Province of residence

Stats SA used geographical stratification for all the household surveys considered in this paper. Until March 2004 the sample was explicitly stratified by province and the urban/rural classification, and since September 2004 explicit stratification has occurred by district council / metro areas. It is therefore not surprising to observe that the weighted shares of the population residing in each province are quite stable across time (Table 4). The one exception is OHS1993, for which the TBVC states were excluded from the sample. As a result, this survey has a much smaller total population (after taking probability weights into consideration), and the provinces that now contain former homelands, such as North West, are underrepresented compared to their share in subsequent years. Given the stark differences in living standards and demographic characteristics between the TBVC states and the rest of South Africa, those living in the TBVC states can by no means be considered to be a random sample from the total population, so that OHS1993 is not comparable to subsequent surveys.

Table 4: Provincial share of population, all ages, 1993 – 2006

	WC	EC	NC	FS	KZN	NW	GAU	MPU	LIM	Population
OHS1993	11.5%	6.0%	2.2%	7.4%	26.1%	2.9%	23.2%	6.7%	14.0%	32,207,758
OHS1994	9.1%	15.9%	1.8%	6.4%	21.3%	8.0%	17.1%	7.2%	13.1%	40,251,142
OHS1995	9.8%	15.6%	2.1%	6.5%	20.8%	8.3%	18.1%	6.9%	12.2%	39,659,831
OHS1996	9.8%	15.5%	2.1%	6.5%	20.7%	8.3%	18.1%	6.9%	12.1%	40,582,538
OHS1997	9.7%	15.5%	2.1%	6.5%	20.7%	8.3%	18.1%	6.9%	12.1%	41,443,101
OHS1998	9.7%	15.6%	2.1%	6.5%	20.7%	8.3%	18.0%	6.9%	12.3%	42,235,733
OHS1999	9.6%	15.6%	2.1%	6.5%	20.8%	8.3%	18.0%	6.9%	12.3%	43,271,686
LFS2000a	9.6%	15.6%	2.1%	6.5%	20.8%	8.3%	17.9%	6.9%	12.3%	43,620,361
LFS2000b	9.4%	15.5%	1.9%	6.5%	20.7%	8.2%	18.5%	6.9%	12.2%	44,821,345
LFS2001a	9.4%	15.5%	1.9%	6.5%	20.7%	8.2%	18.5%	6.9%	12.2%	45,080,410
LFS2001b	9.4%	15.5%	1.9%	6.5%	20.7%	8.2%	18.5%	6.9%	12.2%	45,081,045
LFS2002a	9.5%	15.4%	1.9%	6.4%	20.7%	8.2%	18.6%	6.9%	12.2%	45,324,735
LFS2002b	9.6%	15.4%	1.9%	6.4%	20.7%	8.2%	18.7%	6.9%	12.2%	45,560,990
LFS2003a	9.6%	15.3%	1.9%	6.4%	20.7%	8.2%	18.8%	6.9%	12.1%	45,810,074
LFS2003b	9.7%	15.2%	1.9%	6.4%	20.7%	8.2%	18.9%	6.9%	12.1%	46,046,026
LFS2004a	9.7%	15.2%	1.9%	6.4%	20.6%	8.2%	19.0%	6.9%	12.1%	46,270,894
LFS2004b	9.8%	15.1%	1.9%	6.3%	20.6%	8.1%	19.2%	6.9%	12.0%	46,490,122
LFS2005a	9.9%	15.1%	1.9%	6.3%	20.6%	8.2%	19.2%	6.9%	12.0%	46,699,967
LFS2005b	9.9%	15.0%	1.9%	6.3%	20.6%	8.1%	19.3%	6.9%	12.0%	46,917,195
LFS2006a	10.0%	14.9%	1.9%	6.3%	20.5%	8.1%	19.4%	6.9%	12.0%	47,184,311
LFS2006b	10.0%	14.5%	2.3%	6.2%	20.9%	7.1%	20.1%	7.4%	11.3%	47,429,106

Table 4 also indicates that the share of the population residing in Western Cape and Gauteng experienced a sudden increase of approximately one percentage point each between 1994 and 1995, at the cost of the shares of KwaZulu-Natal and Limpopo. It therefore seems as if the population share of the more urbanized provinces were underweighted in OHS1994. On the other hand, the provincial shares in LFS2006b seem to be inconsistent compared with the other LFSs, since the Northern Cape, KwaZulu-Natal, Gauteng and Mpumalanga shares increased abruptly by at least 0.4 percentage points each, at the cost of the dwindling shares of Eastern Cape, North West and Limpopo.

Table 5 shows that if only the population of working age is considered, then provincial shares were much more consistent over the years under investigation, except OHS1993. Specifically, the small inconsistency between the OHS1994 and OHS1995 almost completely disappeared, while the discrepancy problem in LFS2006b, despite still being present, has become less serious.

Finally, looking at the growth of the provincial shares, regardless of whether one looks at the total population or working age population, there is a small increase in the share of the population living in Gauteng and, to a lesser extent, Western Cape. The increases occurring in the latter provinces happen at the cost of a declining share of the population residing in Eastern Cape.

Table 5: Provincial share of working age population, 1993 – 2006

	WC	**EC**	**NC**	**FS**	**KZN**	**NW**	**GAU**	**MPU**	**LIM**	**Population**
OHS1993	12.7%	6.3%	2.2%	7.3%	24.5%	3.0%	26.4%	6.4%	11.2%	19,626,697
OHS1994	10.4%	14.4%	1.9%	6.5%	20.5%	8.0%	20.6%	7.0%	10.7%	24,074,568
OHS1995	10.5%	14.0%	2.1%	6.8%	20.3%	8.3%	20.9%	6.7%	10.4%	24,190,583
OHS1996	10.5%	14.0%	2.1%	6.8%	20.3%	8.3%	20.9%	6.7%	10.3%	24,909,065
OHS1997	10.5%	14.0%	2.1%	6.8%	20.3%	8.3%	20.9%	6.7%	10.3%	25,506,089
OHS1998	10.4%	14.0%	2.1%	6.9%	20.3%	8.4%	20.9%	6.8%	10.5%	25,665,233
OHS1999	10.4%	14.0%	2.1%	6.8%	20.3%	8.3%	20.8%	6.8%	10.5%	26,246,545
LFS2000a	10.4%	14.0%	2.1%	6.8%	20.3%	8.3%	20.8%	6.7%	10.5%	26,465,110
LFS2000b	10.2%	13.9%	2.0%	7.0%	20.2%	8.3%	21.0%	6.8%	10.7%	27,836,456
LFS2001a	10.3%	13.9%	2.0%	6.9%	20.2%	8.3%	21.1%	6.8%	10.6%	28,062,004
LFS2001b	10.2%	13.8%	2.0%	6.9%	20.6%	8.3%	20.9%	6.7%	10.8%	28,084,327
LFS2002a	10.3%	13.9%	2.0%	6.8%	20.4%	8.2%	21.3%	6.7%	10.6%	28,298,255
LFS2002b	10.5%	13.9%	2.0%	6.7%	20.1%	8.2%	21.3%	6.8%	10.7%	28,495,088
LFS2003a	10.5%	13.7%	2.0%	6.7%	20.4%	8.3%	21.3%	6.7%	10.6%	28,724,521
LFS2003b	10.5%	13.9%	1.9%	6.6%	20.4%	8.3%	21.1%	6.7%	10.6%	28,906,230
LFS2004a	10.7%	13.7%	1.9%	6.7%	20.2%	8.2%	21.3%	6.7%	10.6%	29,099,787
LFS2004b	10.7%	13.7%	1.9%	6.5%	20.2%	8.2%	21.4%	6.6%	10.7%	29,270,821
LFS2005a	10.7%	13.7%	1.9%	6.5%	20.3%	8.2%	21.5%	6.6%	10.6%	29,489,763
LFS2005b	10.7%	13.4%	1.9%	6.4%	20.4%	8.1%	21.7%	6.6%	10.7%	29,663,379
LFS2006a	10.8%	13.4%	1.9%	6.4%	20.3%	8.1%	21.7%	6.7%	10.6%	29,817,824
LFS2006b	10.7%	13.1%	2.4%	6.4%	20.4%	7.2%	22.5%	7.3%	10.1%	29,972,521

3.1.2 Area type

In the OHS1993 and OHS1994 datasets, there is a dummy variable indicating whether each household resided in an urban or a rural, but there is no explanation on how the variable is derived. OHS1995 identifies areas as urban, semi-urban and rural[12]. It was considered best for comparison purposes to collapse these three categories into a dichotomous urban-rural variable by reclassifying all semi-urban areas as rural[13].

In OHS1996 and OHS1997, a rural-urban dummy variable was again provided without an explanation, and was presumably derived from the stratification (enumerator areas). The metadata files for OHS1998 and OHS1999 state that urban areas refer to "enumerator areas within municipal or local authority boundaries", whereas rural areas are "enumerator areas with population concentrations adjacent to a municipal border (an enumerator area must have one common boundary with the municipal border) and enumerator areas situated in rural areas (not sharing a common boundary with a proclaimed urban municipal area)."

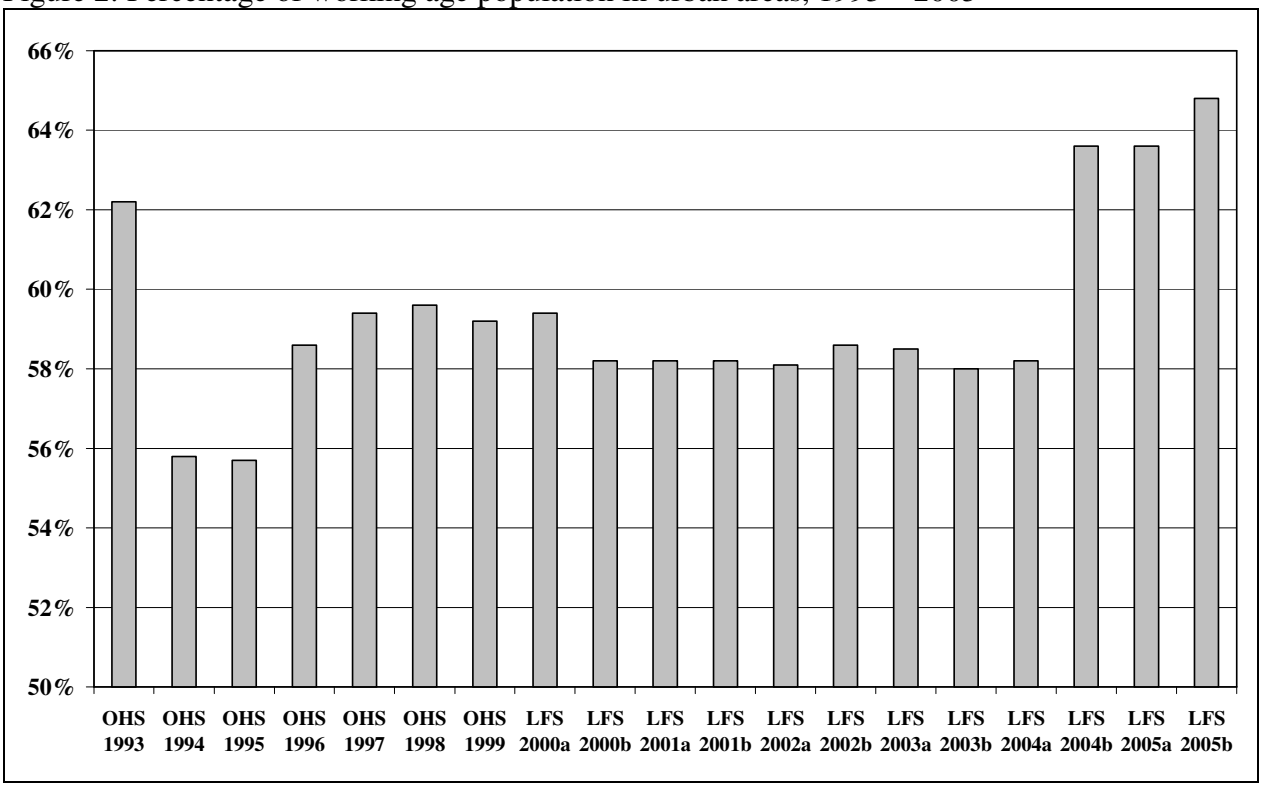
From LFS2000a to LFS2004a, the urban-rural variable can be derived from the stratum variable, which is used to stratify the sample into 18 strata: rural and urban areas for each of the nine provinces. Starting in LFS2004b, the stratum variable is no longer provided, since stratification now occurred along district councils, but it is possible to use the primary sampling unit data to classify the areas into "formal urban", "informal urban", "formal rural" and "tribal areas"., where the first two correspond to urban areas (personal correspondence with Stats SA). In 2006, the stratification changed again, so that none of the afore-mentioned strategies can be used to create an urban-rural variable.

It can be observed from Figure 2 that there was a 7 percentage point decrease in the share of the working age population living in urban areas between 1993 and 1994 (which is caused by the exclusion of the TBVC states in 1993), followed by a three percentage point increase between 1995 and 1996. Between OHS1996 and LFS2004a the urban share stabilized at approximately 58%-59%, while the new sampling methodology results in an over-estimation of this share by more than five percentage points between LFS2004a and LFS2004b.

[12] According to the 1995 OHS metadata file, urban areas include "Ordinary town or city area[s] as well as vacant areas within municipal boundaries", "Area[s] with informal dwellings (e.g. squatter camp)", "Area[s] with mainly hostels, (e.g. mine, factory and municipal hostels)" and "Area[s] with mainly hospital and prison institutions within municipal/local authority boundaries". Semi-urban areas are "Semi-towns (i.e. a town without a local authority) with predominantly formal dwellings", "Area[s] with informal dwellings (e.g. squatter camp)", "Area[s] in which mainly hostels are found", and "Area[s] with mainly hospital and prison institutions". Finally, rural areas consist of "Semi-towns (i.e. a town without a local authority) with predominantly formal dwellings", Town"[s] ("village") without a local authority and which is not situated within a tribal area and with formal and semi-formal dwellings such as houses, huts and rondavels", "Villages/settlements within a tribal area", "Area[s] with population concentration in informal dwellings (e.g. squatter camp)", "Area[s] with mainly hostels where housing for employees is provided by employers (such as mines, factories and power stations)", "Area[s] with mainly hospital and prison institutions", "Area[s] with farms, agricultural holdings, holiday resorts, agricultural schools and colleges and other rural area", as well as "Tribal areas excluding villages/settlements".

[13] In OHS1995, only 1.8% of the population resided in semi-urban areas.

Figure 2: Percentage of working age population in urban areas, 1993 – 2005



3.1.3 Race

The surveys differentiate the four main South African population groups: "Black" "Coloured", "Asian" and "White". Prior to 1995, the survey questionnaires identified these groups using the above-mentioned denominations, but in 1995 (and in all the surveys since) the Black population category was relabeled "Black/African" and the Indian population category was changed to "Indian/Asian". OHS1997 and OHS1998 also allowed respondents to identify themselves as "Griqua", but only a very small proportion of the population did – never more than a twentieth of a percentage point of the population – so this option was removed again for OHS1999. In comparing different surveys it would make sense to either omit these individuals, or to reclassify them as "Coloured" (or maybe as "unspecified" or "other").

Starting in OHS1997, the surveys allowed individuals to be classified as belonging to an "Other" population group (in addition to the aforementioned race groups). Since there is no indication of which groups this could refer to, these individuals are best grouped together with those who failed to specify any race group. Prior to 1998, no individuals were classified as belonging to an "Unspecified" race group, but since 1998 a small proportion of people fell into this category. In terms of their labour market outcomes (specifically, the average unemployment rate and earnings level) this group falls somewhere between Whites and Indians, suggesting that it is predominantly Whites who refuse to specify their race. It is unclear whether the fact that no individuals were classified as belonging to an "unspecified" race prior to 1998 means that all respondents answered this question, whether these values were somehow inferred or whether these individuals were excluded from the sample.

The omission of the TBVC states in 1993 naturally resulted in a lower share of Black individuals than in the other survey years, but Table 6 shows that this appears to have been only partly corrected in OHS1994. Comparing 1994 to any of the later survey years, one needs to keep in mind that this survey suffered from an under-sampling of Blacks and an over-sampling of Whites (the other two population groups appear to have been sampled correctly). Table 6 also shows that for the remaining surveys, the data shows fairly smooth trends in racial shares, with Blacks increasing their share in the working age population whilst the other three population groups are experiencing decreasing shares.

Table 6: Racial share of working age population, 1993 – 2006

	Black	**Coloured**	**Indian**	**White**	**Unspecified**
OHS1993	67.3%	11.0%	3.5%	18.2%	0.0%
OHS1994	72.8%	9.3%	2.9%	15.0%	0.0%
OHS1995	75.1%	9.4%	2.9%	12.6%	0.0%
OHS1996	75.2%	9.4%	2.9%	12.5%	0.0%
OHS1997	75.5%	9.4%	2.9%	12.2%	0.0%
OHS1998	75.6%	9.3%	2.9%	12.1%	0.1%
OHS1999	75.9%	9.2%	2.9%	11.8%	0.1%
LFS2000a	75.7%	9.3%	2.9%	12.1%	0.0%
LFS2000b	76.2%	9.1%	2.8%	11.7%	0.2%
LFS2001a	76.4%	9.1%	2.8%	11.5%	0.1%
LFS2001b	76.4%	9.0%	3.0%	11.5%	0.1%
LFS2002a	76.4%	9.1%	3.0%	11.4%	0.1%
LFS2002b	76.6%	9.1%	2.9%	11.2%	0.1%
LFS2003a	77.0%	9.1%	3.0%	10.9%	0.1%
LFS2003b	77.0%	9.1%	2.9%	11.0%	0.1%
LFS2004a	77.2%	9.1%	2.9%	10.8%	0.0%
LFS2004b	77.3%	9.2%	2.8%	10.6%	0.2%
LFS2005a	77.4%	9.1%	2.8%	10.5%	0.1%
LFS2005b	77.5%	9.0%	2.8%	10.4%	0.2%
LFS2006a	77.6%	9.0%	2.8%	10.5%	0.1%
LFS2006b	79.3%	8.8%	2.5%	9.2%	0.3%

3.1.4 Gender

It can be observed from Table 7 that between 1993 and 2006, the share of males amongst working age individuals mostly fluctuated between 47% and 49%, with only a negligible number being coded as unspecified. There appears to have been a slight oversampling of men in 1993 and to an even lesser extent in 1994, whereas women were slightly over-sampled in 1996. OHS1993 may well have captured the (predominantly male) migrant workers of households living in the TBVC states, whilst failing to capture the rest of the households.

Table 7: Gender share of working age population, 1993 – 2006

	Male	**Female**	**Unspecified**
OHS1993	49.0%	51.0%	0.0%
OHS1994	48.8%	51.2%	0.0%
OHS1995	47.7%	52.3%	0.0%
OHS1996	47.0%	53.0%	0.0%
OHS1997	47.9%	52.1%	0.0%
OHS1998	47.9%	52.1%	0.0%
OHS1999	48.0%	52.0%	0.1%
LFS2000a	47.7%	52.3%	0.0%
LFS2000b	48.4%	51.5%	0.0%
LFS2001a	48.6%	51.4%	0.0%
LFS2001b	48.4%	51.6%	0.0%
LFS2002a	48.3%	51.6%	0.0%
LFS2002b	48.7%	51.2%	0.0%
LFS2003a	48.6%	51.4%	0.0%
LFS2003b	48.4%	51.6%	0.0%
LFS2004a	48.3%	51.6%	0.0%
LFS2004b	48.4%	51.5%	0.0%
LFS2005a	48.2%	51.7%	0.1%
LFS2005b	48.1%	51.8%	0.1%
LFS2006a	48.3%	51.7%	0.0%
LFS2006b	49.2%	50.8%	0.0%

3.1.5 Marital Status

From Figure 3, it can be seen that since 1993 the percentages of the working age population that were married or living with their partners have hovered around 40%, and probably displays a downward trend over time. OHS1998 was clearly an outlier: for this year the percentage fell from 41.7% to 20.6%, and then returned to a rate exceeding 40.5% in OHS1999.

Figure 3: Percentage of working age population who are married or living together with a partner as husband and wife, 1993 – 2006

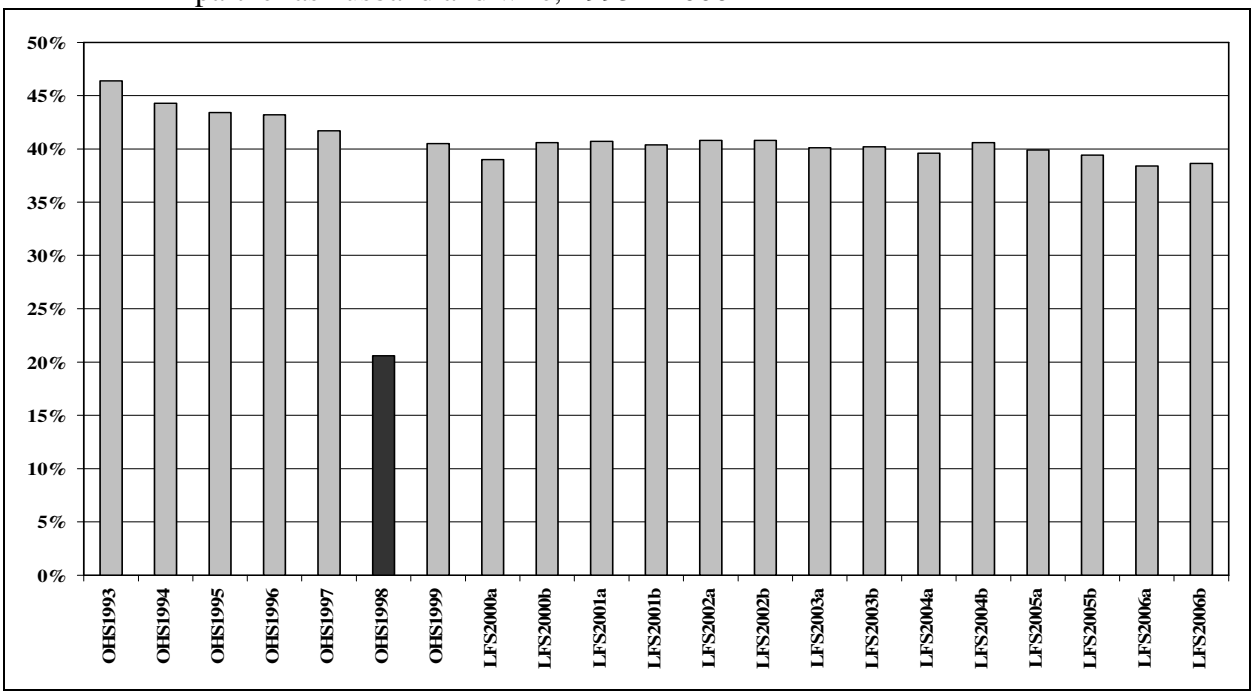


Table 8 below shows the percentage of the working age population in each marital status category in the OHS1997 and OHS1998. As may be seen in the table, more than half of the working age population (50.5%) is reported to be divorced in 1998. The most sensible explanation is that the categories were miscoded. In fact, if the 1998 values are recoded according to the codes used in OHS1999, the shares become very similar to those in the OHS1997 and OHS1999. The share of divorcees in OHS1998 is now 2.9%, up slightly from 2.6% in the previous year, and share of people that are married or living together now stands at a much more credible 42.4%.

Table 8: Percentage of the working age population in each marital status category, OHS1997 – OHS1999

Code	Meaning	OHS1997	OHS1998	OHS1998 (corrected)	OHS1999
1	Single	53.0%	25.2%	50.5%	51.4%
2	Married – civil	25.6%	11.2%	25.2%	25.1%
3	Married – traditional	9.7%	6.0%	11.2%	10.9%
4	Live together	4.8%	4.1%	6.0%	6.3%
5	Widow	4.2%	2.9%	4.1%	3.6%
6	Divorced	2.6%	50.5%	2.9%	2.8%

Table A.2 in the appendix shows the corrected figures for each category of marital status for all the surveys. For the first nine LFSs, the categories "married" and "living together"[14] were lumped together, so that in order to validly compare data from these surveys to those preceding and succeeding it, these two categories must be added together for the other years as well.

Table 9 shows that the share of the working age population who were "married or living together" decreased substantially – maybe unrealistically so – between 1993 and 1995 from 46.4% to 43.4%, after which the decline continued more steadily until 2006 (reaching 38.4%). Using LFSs after March 2004, this decline resulted wholly from the decrease in the share of working aged individuals who were married, and occurred despite an increase in the share of people living together (See Table A.2). The percentage of people who have never been married increased steadily over the period, combined with a small increase in widows and widowers. The share of divorcees remained more or less stable. Although the unspecified category is rather small, it implies that including a variable for marital status in an employment or earnings regression will result in the loss of almost a percentage point of the total sample of the working age population.

Table 9: Marital status in South Africa, 1993 – 2006

	Never married	Married or live together*	Widow/ Widower	Divorced	Unspecified	Total
OHS1993	47.60%	46.40%	3.52%	2.49%	0.00%	100.00%
OHS1994	49.67%	44.27%	3.46%	2.60%	0.00%	100.00%
OHS1995	50.61%	43.39%	3.53%	2.46%	0.00%	100.00%
OHS1996	50.32%	43.18%	3.62%	2.46%	0.41%	100.00%
OHS1997	52.00%	41.70%	3.60%	2.70%	0.00%	100.00%
OHS1998	51.22%	42.13%	3.56%	2.82%	0.27%	100.00%
OHS1999	52.57%	40.50%	3.26%	3.03%	0.64%	100.00%
LFS2000a	53.75%	39.00%	3.73%	3.45%	0.06%	100.00%
LFS2000b	52.05%	40.57%	4.19%	3.05%	0.14%	100.00%
LFS2001a	52.02%	40.74%	4.21%	3.01%	0.02%	100.00%
LFS2001b	52.56%	40.41%	4.12%	2.90%	0.01%	100.00%
LFS2002a	52.18%	40.78%	4.04%	2.99%	0.01%	100.00%
LFS2002b	52.24%	40.78%	3.94%	3.04%	0.01%	100.00%
LFS2003a	52.87%	40.14%	4.14%	2.81%	0.03%	100.00%
LFS2003b	52.66%	40.16%	4.20%	2.96%	0.02%	100.00%
LFS2004a	53.40%	39.64%	4.04%	2.88%	0.03%	100.00%
LFS2004b	52.33%	40.63%	4.10%	2.94%	0.01%	100.00%
LFS2005a	53.41%	39.92%	3.90%	2.74%	0.04%	100.00%
LFS2005b	53.89%	39.41%	3.99%	2.70%	0.01%	100.00%
LFS2006a	55.06%	38.40%	3.98%	2.56%	0.00%	100.00%
LFS2006b	54.71%	38.62%	4.22%	2.44%	0.00%	100.00%

* Married or live together = 'married – civil' + 'married – traditional' + 'live together' from OHS1994 to OHS1999.

* Married or live together = 'married' + 'live together' in OHS1993 and from LFS2004b to LFS2006b.

[14] Between 1993 and 1995, this category only included those who categorised themselves as "living together". For the 1996 to 1999 surveys, this category was relabeled "living together with a partner", before being absorbed into the "married or living together as husband and wife" category for the first LFSs. Starting in 2004b, the questionnaires started distinguishing between "married" and "living together as husband and wife". There are no observable discontinuities in the shares of the population who classify themselves as living together that correspond to the changing definition, however. Similarly, the proportion of individuals who classified themselves a "divorced" in 1993 does not differ much from those who classified themselves as "divorced / separated" after the re-labeling of this category starting in 1994.

3.2 EDUCATIONAL ATTAINMENT VARIABLES

3.2.1 Highest educational attainment

The surveys asked respondents to give "the highest level of schooling completed". The question was asked in a fairly consistent manner across surveys, although OHS1997 and OHS1998 used two questions: the first asked "What is the highest school class/standard that (the person) completed", whereas the second read "Does (the person) have a technical or artisan certificate, diploma or degree, completed at an educational institution?" Tables A.3 to A.8 in the appendix provide a more in-depth explanation on the educational attainment categories in each survey, as well as how the years of education (eduyear) are derived. This variable was then used to derive the seven broad attainment categories for Table 10.

Table 10 shows that the share of the population with no education or primary education decreased substantially over the period, whereas the proportion with incomplete secondary education remained fairly stable, and the share with completed secondary (matriculation) or some form of post-secondary qualification increased. It can also be seen from the table that the percentage of the population with post-Matric certificates or diplomas was slightly higher in OHS1995 than in the preceding and subsequent years. Besides, OHS1994 showed an over-sampling of people with primary or incomplete secondary education, whilst under-sampling those with completed secondary education.

Table 10: Percentage of working age population in each broad educational attainment category, 1993 – 2006

	No Schooling	**Primary**	**Incomplete secondary**	**Matric**	**Matric + Certificate / Diploma**	**Degree**	**Unspecified/ Don't know/ Others**
OHS1993	10.4%	25.4%	40.5%	16.9%	3.3%	3.1%	0.4%
OHS1994	9.2%	26.7%	41.5%	15.7%	4.2%	2.4%	0.4%
OHS1995	9.4%	24.7%	40.7%	17.1%	5.0%	2.3%	0.9%
OHS1996	10.1%	25.4%	39.1%	17.5%	4.1%	2.5%	1.3%
OHS1997	9.9%	25.4%	40.7%	17.2%	4.3%	2.2%	0.2%
OHS1998	9.9%	25.9%	39.8%	17.7%	4.6%	2.1%	0.2%
OHS1999	8.2%	26.9%	38.1%	18.3%	3.9%	3.2%	1.5%
LFS2000a	7.8%	26.0%	39.6%	18.4%	4.5%	2.9%	1.0%
LFS2000b	7.9%	26.9%	38.7%	17.3%	4.7%	3.6%	0.9%
LFS2001a	7.6%	25.9%	39.0%	19.0%	4.6%	3.2%	0.7%
LFS2001b	7.8%	26.5%	38.3%	18.9%	4.4%	3.2%	0.9%
LFS2002a	7.6%	24.6%	39.6%	19.8%	4.5%	3.2%	0.7%
LFS2002b	7.3%	24.8%	39.8%	19.5%	4.6%	3.3%	0.8%
LFS2003a	7.2%	24.0%	40.2%	20.3%	4.5%	3.3%	0.6%
LFS2003b	6.9%	23.9%	39.9%	21.0%	4.6%	3.2%	0.5%
LFS2004a	7.0%	23.3%	39.9%	21.8%	4.4%	3.3%	0.4%
LFS2004b	7.0%	22.8%	40.7%	21.3%	4.4%	3.1%	0.8%
LFS2005a	6.5%	22.0%	41.1%	22.1%	4.6%	3.2%	0.5%
LFS2005b	6.6%	22.0%	41.0%	21.9%	4.8%	3.2%	0.6%
LFS2006a	6.2%	21.0%	41.2%	22.9%	5.2%	3.2%	0.4%
LFS2006b	6.0%	20.9%	41.9%	22.4%	5.2%	3.0%	0.5%

3.2.2 Attendance at educational institutions at the time of the survey

The surveys also asked respondents whether they were attending an educational institution at the time of the survey and Table A.9 provides more information on how this question was asked in each survey. Interestingly, the answers given to this question sometimes appeared to contradict the responses to the educational attainment question. Table 11 presents the exact numbers of respondents whose answers to these two questions were inconsistent, presented in three separate columns: those who reported not having completed secondary education but were attending university or a technikon at the time of the survey; those who reported having completed some form of a post-secondary qualification, but were attending school or pre-school at the time of the survey; and those who reported having Matric as their highest level of completed education but were attending school at the time of the survey.

It is possible for someone who completed her secondary education to repeat Matric, so that not everyone in the final column of Table 11 necessarily represent inconsistent responses. However, the number of such responses suggests that something peculiar was taking place. One possibility is that individuals who wrote Matric but failed often responded that their highest level of completed education is completed secondary. If many of those in Matric were repeating, then this could explain the high number of individuals who answer inconsistently in this manner. Unfortunately, this will also cast some doubt over the reliability of all those who reported that they have indeed completed secondary education, but who were no longer in the school system.

Note that none of these inconsistencies occurred from OHS1993 to OHS1998. Again, it is not known whether this is due to the fieldworkers not allowing inconsistent responses, or whether Stats SA somehow corrected inconsistencies prior to releasing the data, or whether these individuals were simply omitted from the final sample.

Table 11: Number of respondents in the working age population giving inconsistent answers regarding educational attainment (unweighted), 1999 – 2006

	A	**B**	**C**	**D**	**(A + B + C) / D**
OHS1999	112	35	519	14776	4.5%
LFS2000a	37	1	124	5033	3.2%
LFS2000b	88	25	559	14264	4.7%
LFS2001a	95	13	429	13804	3.9%
LFS2001b	106	8	320	13559	3.2%
LFS2002a	109	14	272	13590	2.9%
LFS2002b	86	6	200	13098	2.2%
LFS2003a	65	13	242	12820	2.5%
LFS2003b	82	11	180	12521	2.2%
LFS2004a	95	5	167	12324	2.2%
LFS2004b	61	9	228	13992	2.1%
LFS2005a	82	6	284	14034	2.7%
LFS2005b	74	10	357	14161	3.1%
LFS2006a	56	4	326	13914	2.8%
LFS2006b	69	8	264	13883	2.5%

Note: the number of people in A, B and C is zero from OHS1993 to OHS1998.
A: Incomplete secondary education, but attending university / technikon at the time of survey
B: Post-secondary education, but attending school at the time of survey
C: Completed secondary education, but attending school at the time of survey
D: Number of people attending educational institutions at the time of the survey

3.3 EMPLOYMENT VARIABLES[15]

3.3.1 Number of employed

The employed refers to all the respondents who reported engaging in employment activities in the week preceding the interview. Although the survey questions used to identify the employed were not asked in a consistent manner[16], there has been an increasing effort to capture low-income employment since the earliest household surveys, so that the increase in employment revealed by a comparison of surveys is at least partly artificial. Examining the first question asked to determine the labour market status of respondents is informative for illustrating this trend.

In OHS1995, the respondents were asked "[w]hat [they] did most during the last 7 days", whereas in OHS1996 respondents were asked whether "During the past 7 days [they] actually [did] work for pay, profit or family gain?". Note that "working full-time" and "working part-time" were allowed as answers. The 1997 and 1998 surveys added a third alternative to the questionnaire, "casual work", and in 1999, "seasonal worker" was also added as an option. From 2000 onwards, respondents were asked if they engaged in any one of numerous specific (and mostly low-income) activities such as "guarding cars" or "making things for sale", for example (Burger and Yu, 2006: 5). It is therefore clear that increased effort was made to capture informal and low-income employment and an implication of this increased effort may well be apparent fluctuations in the number of employed between surveys.

Table 12: Number of employed, 1995 – 2006

	Black	Coloured	Indian	White	Total*	% change of total employed
OHS1995	6,136,137	1,144,836	358,589	1,859,785	9,499,347	
OHS1996	5,489,346	1,222,031	337,118	1,917,812	8,966,307	-5.6%
OHS1997	5,713,778	1,161,019	361,837	1,857,013	9,093,647	1.4%
OHS1998	5,915,277	1,168,302	342,141	1,934,031	9,370,130	3.0%
OHS1999	6,659,911	1,285,810	391,951	2,001,963	10,356,143	10.5%
LFS2000a	8,120,175	1,317,383	394,599	2,035,873	11,874,409	14.7%
LFS2000b	8,363,113	1,332,926	407,860	2,095,919	12,224,406	2.9%
LFS2001a	8,455,545	1,320,941	409,630	2,055,501	12,260,207	0.3%
LFS2001b	7,344,392	1,277,194	428,345	2,099,927	11,167,541	-8.9%
LFS2002a	7,776,952	1,311,916	406,219	2,092,780	11,603,398	3.9%
LFS2002b	7,506,688	1,292,001	429,390	2,042,567	11,283,924	-2.8%
LFS2003a	7,497,609	1,337,553	411,287	2,041,843	11,297,621	0.1%
LFS2003b	7,570,529	1,309,498	432,700	2,090,445	11,411,351	1.0%
LFS2004a	7,540,422	1,388,152	420,024	2,022,965	11,378,217	-0.3%
LFS2004b	7,866,030	1,296,317	418,797	2,014,698	11,630,196	2.2%
LFS2005a	8,079,850	1,356,286	422,606	2,011,964	11,894,320	2.3%
LFS2005b	8,497,599	1,327,511	440,182	1,991,480	12,287,798	3.3%
LFS2006a	8,567,842	1,387,420	429,705	2,036,940	12,437,963	1.2%
LFS2006b	8,873,535	1,410,063	451,410	2,005,587	12,787,285	2.9%

**** Including unspecified race groups***

[15] For the remainder of Section 3, the focus will be on the working age population from OHS1995 to LFS2006b.

[16] Figures A.1 to A.10 provides detailed explanation on the whole algorithm for identifying the employed, while Tables A.10 and A.11 provide information on the people who are qualified as employed immediately under the algorithm. Unfortunately, the Stats SA metadata did not provide any explanation on how the employment status is derived in OHS1995.

Table 12 presents that number of employed for all the years under investigation, as well as the percentage change in the number of employed between surveys. The latter appears to be relatively stable with the exception of the much larger decreases between OHS1995 and OHS1996 (-5.6%) and between LFS2001a and LFS2001b (-8.9%), as well as the large increases between OHS1998 and OHS1999 (10.5%), and between LFS2000a and LFS2000b (14.7%).

On the other hand, Table 13 below presents a break down of the number of employed by the type of employment. From the table, it can be seen that self-employment is not well-captured in the OHS years since the self-employed as a percentage of all the employed is lower than it is in the LFS years. The table also indicates that, although there is only a small decrease in the number of employees in LFS2000a, employees as percentage of total employed drops substantially (from 85.5% in OHS1999 to 74.1% in LFS2000a). This is caused by the serious over-estimation of the number of self-employed in LFS2000a.

Employees as percentage of all employed remains at a low level in LFS2000b and LFS2001a. From LFS2001b onwards, this share stabilizes at approximately 80% but still fails to return to the high levels reported between OHS1995 and OHS1999.

Table 13: Type of employment, 1995 – 2006

	Employee		**Self-Employed**		**Unspecified**	**Total Employed**
	Number	**Percentage**	**Number**	**Percentage**		
OHS1995	8,123,412	85.5%	1,375,935	14.5%	0	9,499,347
OHS1996	8,313,240	93.2%	611,045	6.8%	42,022	8,966,307
OHS1997	8,167,479	89.8%	926,168	10.2%	0	9,093,647
OHS1998	8,339,925	89.0%	1,025,748	11.0%	4,457	9,370,130
OHS1999	8,844,574	85.5%	1,505,706	14.5%	5,863	10,356,143
LFS2000a	8,787,145	74.1%	3,073,630	25.9%	13,634	11,874,409
LFS2000b	9,370,733	76.8%	2,825,474	23.2%	28,199	12,224,406
LFS2001a	9,024,720	73.7%	3,218,407	26.3%	17,080	12,260,207
LFS2001b	9,011,975	80.8%	2,144,102	19.2%	11,464	11,167,541
LFS2002a	9,081,627	78.4%	2,508,940	21.6%	12,831	11,603,398
LFS2002b	9,081,716	80.6%	2,190,994	19.4%	11,214	11,283,924
LFS2003a	9,194,238	81.4%	2,099,251	18.6%	4,132	11,297,621
LFS2003b	9,276,158	81.3%	2,131,304	18.7%	3,889	11,411,351
LFS2004a	9,356,332	82.3%	2,018,613	17.7%	3,272	11,378,217
LFS2004b	9,414,391	81.0%	2,206,814	19.0%	8,991	11,630,196
LFS2005a	9,535,624	80.3%	2,340,253	19.7%	18,443	11,894,320
LFS2005b	9,846,100	80.3%	2,422,542	19.7%	19,156	12,287,798
LFS2006a	9,771,856	78.7%	2,658,832	21.4%	7,275	12,437,963
LFS2006b	10,184,406	79.7%	2,592,531	20.3%	10,348	12,787,285

3.3.2 Formal sector vs. Informal sector

Figure A.11 in the appendix explains how the employed are classified into different categories of formal and informal sector workers, while Table 14 below disaggregates the employed in each survey by the sector in which they are employed. Given that in OHS1995 and OHS1996 only the self-employed were asked to declare the registration status of their business, it is impossible to derive the formal/informal status of employees in these two years.

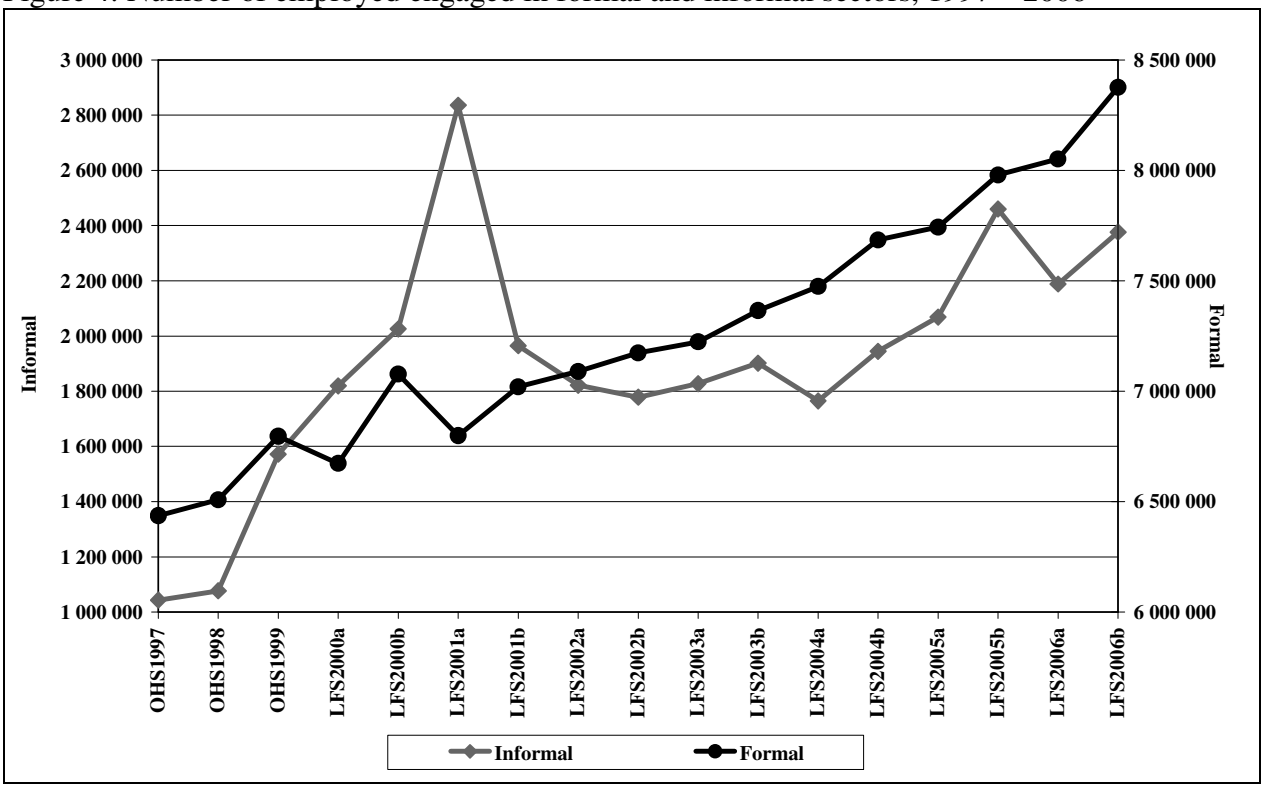
Table 14: Employment by sector, 1995 – 2006

	Domestic Workers	Informal	Formal	Subsistence Agriculture	Commercial Agriculture	Don't know	Unspecified	Total Employed
OHS1995	695,416	521,668	219,213	26,530	49,546	0	7,986,974	9,499,347
OHS1996	766,334	330,100	304,260	24,687	56,296	0	7,484,630	8,966,307
OHS1997	828,254	1,043,347	6,436,017	187,486	525,618	0	72,925	9,093,647
OHS1998	747,281	1,077,141	6,508,097	202,082	725,474	0	110,055	9,370,130
OHS1999	812,465	1,571,646	6,796,008	284,336	798,905	0	92,783	10,356,143
LFS2000a	1,002,719	1,819,556	6,672,951	1,507,625	756,510	86,472	28,576	11,874,409
LFS2000b	941,463	2,026,065	7,077,307	1,074,413	766,917	108,318	229,923	12,224,406
LFS2001a	844,135	2,836,182	6,798,257	742,404	784,712	214,235	40,282	12,260,207
LFS2001b	881,168	1,964,763	7,019,158	382,241	764,521	127,023	28,667	11,167,541
LFS2002a	875,172	1,821,426	7,089,163	862,747	864,576	74,868	15,446	11,603,398
LFS2002b	843,019	1,778,542	7,173,080	550,068	851,897	61,643	25,675	11,283,924
LFS2003a	885,322	1,827,711	7,223,138	443,426	841,440	57,332	19,252	11,297,621
LFS2003b	894,626	1,901,131	7,364,616	365,378	831,526	36,403	17,671	11,411,351
LFS2004a	845,965	1,764,630	7,473,638	340,515	912,831	25,704	14,934	11,378,217
LFS2004b	880,067	1,944,236	7,684,843	425,083	624,358	52,970	18,639	11,630,196
LFS2005a	848,914	2,068,479	7,741,991	513,022	647,448	27,756	46,710	11,894,320
LFS2005b	858,199	2,459,690	7,979,587	337,884	578,059	33,783	40,596	12,287,798
LFS2006a	849,085	2,187,940	8,051,532	702,881	605,795	14,098	26,632	12,437,963
LFS2006b	884,898	2,376,338	8,376,441	472,697	605,129	46,935	24,847	12,787,285

Interestingly, the increase in the number of employed by roughly 3 million (9,093,647 to 12,287,798) between OHS1997 and LFS2005b is partly driven by the informal sector, with nearly half of the new jobs (1.4 million) being created in the latter. This trend suggests that the formal sector is unable to generate sufficient employment opportunities. Burger and Yu (2006: 5) report that informal sector employment increased from roughly 5% of total employment in 1995 to around 14% of total employment in 1997 and 1998, continuing on this upward trend until it stabilized at roughly 21% from 2001 onwards. However, the improvement in the ability of Stats SA in capturing informal, low-income activities may well have played a role in driving this apparent increase in informal employment.

Finally, it appears that informal sector employment was over-estimated in LFS2001a, as well as in LFS2005b (although less so than it is in LFS2001a). Figure 4 presents graphically the trends in both formal and informal sector employment since 1997, clearly showing the over-estimation of informal sector employment in LFS2001a.

Figure 4: Number of employed engaged in formal and informal sectors, 1997 – 2006



3.3.3 Nature of employment

Table 15 below indicates that, with the exception of the sudden decrease in the percentage of the employed engaged in permanent employment in LFS2000b and the relatively smaller percentage of the temporarily employed in OHS1999, the data appear to be consistent. However, there is clearly a downward trend in the percentage of permanently employed. Note that the question on the nature of employment was only asked from OHS1999 onwards.

Table 15: Nature of employment, 1999 – 2006

	Permanent	Fixed period contract	Temporary	Casual	Seasonal
OHS1999	79.2%	2.7%	9.7%	7.1%	1.4%
LFS2000a	78.2%	2.8%	11.4%	6.5%	1.1%
LFS2000b	74.4%	3.8%	12.3%	8.5%	1.0%
LFS2001a	77.5%	3.7%	10.9%	6.8%	1.1%
LFS2001b	77.4%	3.6%	11.4%	6.8%	0.8%
LFS2002a	75.6%	3.1%	13.3%	6.9%	1.1%
LFS2002b	76.0%	4.3%	12.6%	6.4%	0.7%
LFS2003a	75.1%	4.1%	13.2%	6.6%	1.0%
LFS2003b	77.3%	3.8%	11.7%	6.5%	0.7%
LFS2004a	75.7%	3.8%	12.6%	6.7%	1.1%
LFS2004b	75.1%	4.7%	12.8%	6.8%	0.6%
LFS2005a	73.0%	5.3%	12.9%	7.9%	1.0%
LFS2005b	71.7%	5.4%	13.6%	8.6%	0.8%
LFS2006a	71.7%	5.5%	12.2%	9.8%	0.9%
LFS2006b	70.8%	5.7%	13.9%	8.9%	0.7%

3.3.4 Tenure

Tenure stands for the number of years that an employee has worked for his/her present employer. Table 16 indicates inconsistencies in the data on tenure. In some cases, the tenure of the respondent exceeded their age, the tenure reported exceeded (Age – 15) (assuming that respondents started working at age 15), or the tenure was negative. This inconsistency problem was clearly most serious in OHS1995[17], OHS1999[18] and LFS2000b[19]. Note that the response rate to the tenure question is relatively lower from OHS1996 to OHS1998.

Table 16: Number of respondents giving inconsistent answers regarding tenure (unweighted), 1995 – 2006

	A	B	C	D	E		A/D	B/D	C/D	D/E
OHS1995	39	643	783	26,192	26,192		0.1%	2.5%	3.0%	100.0%
OHS1996	6	243	11	11,237	13,510		0.1%	2.2%	0.1%	83.2%
OHS1997	39	561	40	21,954	24,215		0.2%	2.6%	0.2%	90.7%
OHS1998	20	307	6	12,988	15,440		0.2%	2.4%	0.0%	84.1%
OHS1999	25	381	603	21,507	22,174		0.1%	1.8%	2.8%	97.0%
LFS2000a	17	200	7	7,759	7,967		0.2%	2.6%	0.1%	97.4%
LFS2000b	15	365	550	21,448	21,707		0.1%	1.7%	2.6%	98.8%
LFS2001a	10	352	173	21,793	21,999		0.0%	1.6%	0.8%	99.1%
LFS2001b	7	336	20	21,105	21,318		0.0%	1.6%	0.1%	99.0%
LFS2002a	7	328	51	22,307	22,464		0.0%	1.5%	0.2%	99.3%
LFS2002b	8	323	24	20,282	20,491		0.0%	1.6%	0.1%	99.0%
LFS2003a	9	301	46	20,245	20,381		0.0%	1.5%	0.2%	99.3%
LFS2003b	19	311	13	20,023	20,150		0.1%	1.6%	0.1%	99.4%
LFS2004a	10	328	25	20,216	20,262		0.0%	1.6%	0.1%	99.8%
LFS2004b	5	267	10	20,232	20,377		0.0%	1.3%	0.0%	99.3%
LFS2005a	14	266	65	20,773	20,892		0.1%	1.3%	0.3%	99.4%
LFS2005b	12	280	15	20,518	20,678		0.1%	1.4%	0.1%	99.2%
LFS2006a	8	279	53	20,894	21,041		0.0%	1.3%	0.3%	99.3%
LFS2006b	11	260	20	20,766	20,908		0.1%	1.3%	0.1%	99.3%

A: Number of employees whose tenure exceeded their age

B: Number of employees whose tenure exceeded (age – 15)
[Assuming the respondents started working from the age of 15]

C: Number of employees whose tenure was negative (e.g., in LFS2000a, the survey took place in March 2000, but if a person claimed he had worked for the present employer since May 2000, then the tenure will be –2 months)

D: Total number of employees giving answers on tenure, i.e., stating clearly the starting year and month of working for the employer.

E: Total number of employees

[17] OHS1995: all 783 people with negative tenure claimed that they start working for their present employer from June 1999. Unfortunately the metadata is not available, so it is no known if '1999' actually stands for 'unspecified'.

[18] OHS1999: all 603 people with negative tenure claimed that they start working for their present employer from either November or December 1999.

[19] LFS2000b: all 550 people with negative tenure claimed that they start working for their present employer after September 2000 (October 2000: 454 people, November 2000: 95 people, December 2000: 1 person)

Table 17 shows that the tenure data appears to be fairly consistent throughout the years, after dropping people with tenure exceeding age or negative tenure. However, the proportion of the employees who have worked for 0 – 1 year for the same employer is relatively smaller in OHS1995. Besides, there is a slight upward trend in the proportion of employees working for the present employers for less than 2 years in the LFSs.

Table 17: Number of years the employees have been working for their present employers, 1995 – 2007

	0-1 yr	**1-2 yrs**	**2-3 yrs**	**3-5 yrs**	**5-10 yrs**	**10-20 yrs**	**20-30 yrs**	**> 30 yrs**
OHS1995	14.0%	9.4%	7.8%	12.5%	26.0%	22.6%	6.1%	1.6%
OHS1996	19.3%	10.4%	8.5%	11.9%	21.7%	20.5%	6.0%	1.7%
OHS1997	17.9%	11.0%	8.5%	11.7%	21.5%	21.7%	6.3%	1.5%
OHS1998	18.3%	10.9%	8.8%	13.1%	21.6%	19.6%	6.3%	1.5%
OHS1999	20.1%	10.8%	9.1%	12.9%	19.9%	20.4%	5.7%	1.2%
LFS2000a	20.6%	11.0%	8.8%	12.9%	18.3%	19.8%	6.9%	1.8%
LFS2000b	16.1%	12.7%	10.1%	13.6%	18.4%	20.0%	7.1%	2.0%
LFS2001a	19.6%	10.3%	8.8%	13.2%	18.6%	20.4%	7.3%	1.8%
LFS2001b	19.1%	10.4%	9.1%	12.7%	18.8%	20.9%	7.1%	1.9%
LFS2002a	21.1%	9.9%	8.9%	13.8%	18.2%	19.7%	7.0%	1.5%
LFS2002b	21.4%	11.2%	7.7%	14.4%	17.9%	18.3%	7.5%	1.6%
LFS2003a	21.5%	10.2%	8.0%	14.7%	18.9%	18.0%	7.2%	1.7%
LFS2003b	20.6%	10.5%	8.5%	14.8%	19.9%	17.6%	6.6%	1.6%
LFS2004a	21.6%	10.6%	7.2%	14.6%	19.5%	17.7%	7.2%	1.6%
LFS2004b	20.3%	11.3%	8.1%	15.0%	19.2%	17.6%	6.8%	1.7%
LFS2005a	23.9%	11.3%	8.0%	12.7%	18.5%	16.8%	7.0%	1.9%
LFS2005b	22.8%	11.5%	8.4%	12.8%	19.4%	16.9%	6.5%	1.7%
LFS2006a	24.2%	12.1%	7.9%	12.1%	18.5%	16.5%	7.0%	1.6%
LFS2006b	24.7%	11.9%	8.1%	12.3%	18.7%	15.6%	7.1%	1.7%

Note: the employees with tenure exceeding age or negative tenure are excluded.

3.3.5 Firm size

Firm size refers to the number of workers in an organization, and this question was asked for the first time in LFS2000a. In Table 18, it is obvious that the percentage distribution in LFS2004b is inconsistent in comparison with other surveys. The percentage of employed reporting a firm size of 1 person, 20 – 49 persons and 50 or more persons is considerably higher in LFS2004b than it is in the other surveys, and the percentage with 2 – 4 persons, 5 – 9 persons and 10 – 19 persons is considerably lower in LFS2004b than it is in the other surveys.

Table 18: Firm size – all employed, 2000 – 2006

	1 person	**2-4 persons**	**5-9 persons**	**10-19 persons**	**20-49 persons**	**50 or more**
LFS2000a	21.9%	20.1%	9.9%	10.4%	11.5%	26.2%
LFS2000b	20.6%	18.3%	11.8%	12.2%	12.5%	24.7%
LFS2001a	24.4%	16.3%	10.0%	11.4%	13.0%	25.0%
LFS2001b	19.8%	15.1%	10.0%	12.6%	14.2%	28.3%
LFS2002a	20.7%	17.0%	10.3%	12.5%	13.9%	25.6%
LFS2002b	20.0%	13.8%	9.7%	12.2%	14.9%	29.4%
LFS2003a	19.7%	13.6%	9.9%	12.5%	15.7%	28.6%
LFS2003b	19.9%	13.4%	10.8%	13.1%	16.0%	26.8%
LFS2004a	18.6%	12.8%	9.4%	13.4%	16.6%	29.2%
LFS2004b	20.3%	8.7%	8.1%	7.5%	19.9%	35.4%
LFS2005a	19.3%	14.3%	9.7%	12.5%	16.4%	27.9%
LFS2005b	19.7%	14.1%	10.4%	13.7%	15.7%	26.4%
LFS2006a	19.9%	14.9%	10.3%	13.0%	16.4%	25.6%
LFS2006b	19.6%	13.8%	9.5%	13.1%	15.5%	28.6%

An additional peculiarity with the LFS2004b is that less than a quarter of the employed (5,514 out of 25,083) responded clearly on firm size**.** This is strange because the response rate to the question on firm size exceeds 95% in all the other surveys. A further analysis of the answers given by respondents to the question reveals that in LFS2004b, 17,697 of the employed out of the total 25,083 (i.e. 66.6%) reported being unemployed or economically inactive in the firm size question. Table 19 reports this result. Therefore, it appears that a data-inputting error has occurred.

Table 19: The answers of the employed regarding firm size (unweighted), LFS2004b

Code	**Meaning**	**Number of employed**
0	???	501
1	1 worker	1,364
2	2-4 workers	470
3	5-9 workers	485
4	10-19 workers	466
5	20-49 workers	1,095
6	50-59 workers	1,634
7	Don't know	655
8	Non-active/Unemployed	17,697
9	???	716
		Total: 25,083

Note: the firm size question does not have options '0' and '9', but the LFS2004b data show that some people have been coded as falling into these two categories.

3.3.6 Occupation of the employed

The occupation of the employed refers to the specific job that an employed respondent does. There are 11 broad categories of occupations, and Table 20 shows the percentage of employed in each category. There is an abrupt increase in the percentage of professionals in OHS1997 (from 4.1% in the previous survey to 8.8%), complemented by a sudden decrease in the percentage of associate professionals in the same survey (from 13.7% in the previous survey to 8.3%). Furthermore, there is a peculiar increase in the percentage of skilled agricultural workers in LFS2000a (which coincides with an over-estimation of the number of subsistence agriculture workers, as mentioned in section 3.3.2) and, to a lesser extent, in LFS2000b and LFS2002a. Finally, a relatively higher proportion of employed respondents (6.0%) do not give specific answers regarding their occupation in OHS1996.

Table 20: Occupation of the employed, 1995 – 2006

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
OHS1995	5.3%	3.4%	11.2%	11.9%	11.4%	1.2%	11.8%	11.7%	24.7%	7.3%	0.2%
OHS1996	4.9%	4.1%	13.7%	9.7%	11.6%	2.9%	13.0%	8.7%	16.8%	8.6%	6.0%
OHS1997	7.3%	8.8%	8.3%	8.8%	10.3%	3.0%	14.4%	10.3%	16.6%	9.1%	3.1%
OHS1998	7.8%	5.4%	9.6%	10.0%	12.3%	2.4%	14.0%	10.1%	17.8%	8.0%	2.6%
OHS1999	6.6%	5.3%	10.1%	10.3%	11.8%	4.5%	13.1%	10.5%	18.2%	7.9%	1.7%
LFS2000a	5.3%	3.7%	8.9%	8.8%	11.3%	14.0%	12.1%	9.5%	17.7%	8.4%	0.3%
LFS2000b	4.7%	4.8%	9.3%	8.6%	12.0%	9.8%	13.0%	10.0%	19.7%	7.7%	0.5%
LFS2001a	5.2%	3.8%	9.7%	8.7%	13.6%	7.7%	12.7%	9.5%	21.8%	6.9%	0.4%
LFS2001b	5.9%	4.4%	10.5%	9.8%	12.8%	4.7%	13.7%	10.1%	20.1%	7.9%	0.2%
LFS2002a	6.1%	4.1%	10.4%	9.5%	11.4%	9.1%	12.2%	10.0%	19.3%	7.5%	0.4%
LFS2002b	6.5%	4.4%	10.7%	9.8%	11.0%	6.3%	12.9%	10.2%	20.3%	7.5%	0.4%
LFS2003a	6.3%	4.9%	10.0%	9.7%	11.4%	3.8%	12.4%	10.6%	22.6%	7.8%	0.4%
LFS2003b	7.2%	4.8%	10.1%	10.1%	11.9%	3.0%	12.7%	10.0%	22.1%	7.8%	0.2%
LFS2004a	7.3%	4.7%	9.9%	10.3%	11.8%	2.7%	12.4%	10.2%	23.0%	7.4%	0.1%
LFS2004b	7.8%	3.9%	9.9%	10.0%	12.5%	2.8%	13.2%	9.6%	22.5%	7.6%	0.2%
LFS2005a	6.7%	4.5%	9.5%	10.1%	12.3%	3.6%	13.8%	9.9%	22.4%	7.1%	0.2%
LFS2005b	7.0%	4.8%	9.7%	9.7%	13.1%	2.5%	14.2%	9.2%	22.9%	7.0%	0.2%
LFS2006a	6.9%	4.9%	9.5%	9.7%	12.5%	5.2%	13.7%	8.8%	22.0%	6.8%	0.2%
LFS2006b	6.8%	4.7%	9.6%	9.7%	12.8%	3.4%	15.0%	8.7%	22.2%	6.9%	0.1%

Skilled: A: Legislators, senior officials and managers
B: Professionals
C: Technicians and associate professionals
Semi-skilled: 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: I: Elementary occupations
J: Domestic workers
Unspecified: K: Others/Unspecified

3.3.7 Industry of the employed

Table 21 presents the percentage of employed workers working in each industry. From the table, it is noticeable that there is an unusual increase in the percentage of the employed working in agriculture, hunting, forestry and fishing in OHS1995 (13.0%), LFS2000a (19.2%), LFS2000b (15.6%) and LFS2002a (15.0%). Furthermore, a relatively higher proportion of employed respondents in OHS1996 (6.0% versus 1.8% in OHS1995) did not give specific answers about the industries in which they worked.

Table 21: Industry of the employed, 1995 – 2006

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
OHS1995	13.0%	4.6%	15.1%	0.9%	4.7%	17.5%	5.0%	6.1%	22.9%	8.4%	1.8%
OHS1996	8.5%	2.8%	15.4%	1.4%	4.7%	15.3%	5.4%	8.3%	22.5%	9.0%	6.8%
OHS1997	8.3%	4.3%	16.7%	1.3%	5.6%	17.3%	5.8%	8.0%	20.6%	8.3%	3.7%
OHS1998	10.0%	4.6%	14.7%	1.2%	5.8%	19.0%	5.9%	9.1%	19.7%	8.2%	1.7%
OHS1999	10.6%	4.6%	14.5%	0.8%	5.5%	20.1%	5.2%	9.0%	19.1%	9.3%	1.5%
LFS2000a	19.2%	3.9%	12.4%	0.7%	5.0%	20.5%	4.6%	7.1%	16.0%	10.0%	0.6%
LFS2000b	15.6%	4.9%	12.9%	0.8%	5.6%	20.2%	4.8%	8.0%	17.0%	9.4%	0.8%
LFS2001a	12.9%	4.6%	13.2%	0.8%	5.2%	24.9%	4.7%	8.2%	16.4%	8.4%	0.6%
LFS2001b	10.5%	5.0%	14.5%	0.8%	5.7%	22.0%	4.9%	9.3%	17.8%	9.2%	0.4%
LFS2002a	15.0%	4.7%	13.8%	0.7%	5.0%	20.0%	4.9%	8.9%	17.3%	9.3%	0.5%
LFS2002b	12.6%	5.0%	14.5%	0.7%	5.4%	19.4%	5.1%	9.6%	18.1%	9.1%	0.6%
LFS2003a	11.4%	4.9%	14.0%	0.8%	5.3%	20.6%	5.1%	9.2%	18.7%	9.6%	0.4%
LFS2003b	10.6%	4.8%	13.6%	0.8%	5.8%	21.3%	4.7%	9.6%	19.1%	9.4%	0.3%
LFS2004a	11.1%	4.9%	14.0%	0.9%	5.8%	20.7%	5.1%	9.4%	19.0%	9.0%	0.2%
LFS2004b	9.1%	3.5%	14.7%	0.9%	7.1%	21.8%	4.8%	9.9%	18.8%	9.2%	0.2%
LFS2005a	9.8%	3.6%	13.9%	1.1%	6.8%	22.3%	5.0%	9.6%	18.8%	9.0%	0.3%
LFS2005b	7.5%	3.3%	13.9%	0.8%	7.6%	24.6%	5.0%	10.5%	17.8%	8.7%	0.2%
LFS2006a	10.6%	3.2%	13.9%	0.8%	6.9%	24.1%	4.5%	9.6%	17.5%	8.7%	0.2%
LFS2006b	8.5%	3.1%	13.6%	0.9%	8.0%	23.9%	4.8%	10.2%	18.1%	8.7%	0.3%

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

3.3.8 Public sector vs. Private sector

The question asked to respondents in which they have to define whether they work in the public or the private sector is only asked from LFS2000a onwards and Table 22 reports the result. From the table, it appears that the percentage of the employed in the private sector is under-estimated until LFS2001a. Thereafter, the data appears to be consistent.

Table 22: Percentage of employed engaged in public and private sectors, 2000 – 2006

	Public sector	**Government enterprise**	**Private sector**
LFS2000a	18.2%	3.8%	78.1%
LFS2000b	19.1%	3.9%	77.0%
LFS2001a	18.3%	3.5%	78.2%
LFS2001b	14.3%	2.7%	83.0%
LFS2002a	13.4%	2.6%	84.0%
LFS2002b	13.6%	2.3%	84.0%
LFS2003a	14.4%	2.1%	83.5%
LFS2003b	13.7%	2.3%	84.0%
LFS2004a	13.7%	2.5%	83.8%
LFS2004b	13.8%	2.0%	84.2%
LFS2005a	13.7%	2.1%	84.2%
LFS2005b	13.5%	2.3%	84.3%
LFS2006a	12.8%	2.0%	85.3%
LFS2006b	12.6%	2.0%	85.4%

3.3.9 Usual weekly work hours from the main job

Usual weekly work hours stands for the number of hours worked per week and this was only asked from OHS1997 onwards[20]. It is only since LFS2000a that the respondent was asked to include over-time in the usual weekly work hours. Note that there is another question on the work hours, which asks the respondent to declare the work hours from the main job in the 7 days preceding the survey, and it was asked from OHS1995 onwards, but this question will not be considered in the subsequent analysis.

Table 23 reports the percentage of the employed in various categories of usual weekly work hours. It may be seen that a relatively higher proportion of the employed worked less than 20 hours a week in the two LFS2000 surveys. Similarly, a relatively higher proportion of the employed worked longer than 60 hours a week in LFS2005b. Interestingly, 13 employed respondents actually reported more than 168 hours a week (i.e. more than 24 hours a day), which is clearly impossible. Further analysis may be required on the characteristics (such as industry, occupation, race, earnings, etc.) of the employed respondents who reported unreasonably long working hours.

[20] In OHS1996, the question was asked in a slightly different way as the actual (but not usual) weekly work hours.

Table 23: Usual weekly work hours of employed from the main job, 1997 – 2006

	0-20 hrs	21-40 hrs	41-60hrs	61-80hrs	81-100hrs	>100hrs
OHS1997	2.6%	40.9%	44.6%	6.8%	2.3%	0.4%
OHS1998	3.1%	40.6%	42.7%	8.0%	2.7%	0.5%
OHS1999	5.7%	37.4%	45.5%	7.3%	2.6%	0.5%
LFS2000a	10.5%	31.9%	45.0%	8.8%	2.7%	0.6%
LFS2000b	10.2%	31.5%	44.4%	8.3%	3.6%	0.1%
LFS2001a	7.4%	31.8%	45.8%	9.6%	3.7%	0.6%
LFS2001b	6.1%	34.0%	46.9%	8.8%	3.0%	0.4%
LFS2002a	7.4%	33.7%	47.0%	8.2%	3.0%	0.3%
LFS2002b	5.6%	33.1%	49.6%	8.1%	2.8%	0.4%
LFS2003a	5.9%	37.4%	46.4%	7.4%	2.4%	0.3%
LFS2003b	6.5%	37.5%	46.5%	6.7%	2.3%	0.3%
LFS2004a	5.0%	33.6%	51.7%	6.9%	2.5%	0.3%
LFS2004b	5.9%	37.0%	48.0%	6.5%	2.1%	0.2%
LFS2005a	6.4%	35.7%	47.1%	7.3%	3.0%	0.2%
LFS2005b	6.2%	33.0%	44.3%	11.0%	4.5%	0.5%
LFS2006a	7.9%	35.9%	46.1%	6.9%	2.9%	0.4%
LFS2006b	6.3%	41.4%	43.0%	6.7%	2.4%	0.2%

3.4 EARNINGS OF THE EMPLOYED

3.4.1 Declaration of earnings from the main job

Table 24 reports the percentage of the employed with zero income, unspecified income and extremely high real earnings value. Column A indicates that in LFS2000a and LFS2000b, the proportion of the employed reporting zero income is substantially higher than it is in other surveys. This may have the effect of an under-estimation of mean earnings for these two surveys.

Table 24: Proportion of employed declaring zero / unspecified earnings, 1995 – 2006

	A	B	C	D
OHS1995	0.17%	1.14%	0.02%	98.67%
OHS1996	0.06%	5.64%	0.00%	94.30%
OHS1997	0.28%	5.44%	0.00%	94.28%
OHS1998	0.08%	8.63%	0.00%	91.30%
OHS1999	0.42%	12.56%	0.03%	86.98%
LFS2000a	11.39%	10.01%	0.00%	78.60%
LFS2000b	8.00%	4.21%	0.07%	87.72%
LFS2001a	5.90%	6.73%	0.00%	87.37%
LFS2001b	2.98%	6.25%	0.00%	90.77%
LFS2002a	6.83%	6.83%	0.00%	86.33%
LFS2002b	4.08%	7.93%	0.00%	87.99%
LFS2003a	3.28%	7.12%	0.00%	89.60%
LFS2003b	3.24%	8.69%	0.00%	88.07%
LFS2004a	2.48%	6.78%	0.00%	90.74%
LFS2004b	2.98%	9.77%	0.00%	87.25%
LFS2005a	3.91%	7.64%	0.00%	88.45%
LFS2005b	2.84%	6.96%	0.00%	90.20%
LFS2006a	5.10%	5.09%	0.00%	89.80%
LFS2006b	3.50%	7.07%	0.00%	89.43%

A: Employed with zero income
B: Employed with unspecified income
C: Employed with extremely high real monthly earnings value (Outliers = R500000, 2000 prices)
D: Other employed

Table 25 below reports the skills level of work of the employed with zero or unspecified earnings values. From the table, it may be seen that the majority of the employed who report zero income are engaged in semi-skilled work. Further, most of the employed who do not report their earnings are engaged in semi-skilled or skilled work.

Table 25: Skills level of work of the employed with zero or missing earnings values, 1995 – 2006

	Zero earnings				**Missing/Undeclared earnings**		
	Unskilled	**Semi-skilled**	**Skilled**		**Unskilled**	**Semi-skilled**	**Skilled**
OHS1995	27.5%	50.0%	22.5%		28.9%	46.4%	24.7%
OHS1996	0.0%	48.2%	51.8%		16.3%	51.0%	32.7%
OHS1997	24.5%	35.7%	39.8%		14.7%	46.8%	38.5%
OHS1998	20.6%	49.1%	30.4%		17.9%	44.4%	37.7%
OHS1999	4.6%	92.1%	3.3%		16.3%	54.3%	29.4%
LFS2000a	8.6%	90.5%	0.8%		18.7%	51.7%	29.6%
LFS2000b	17.1%	81.9%	1.0%		15.7%	52.9%	31.5%
LFS2001a	11.0%	87.3%	1.7%		13.6%	49.7%	36.7%
LFS2001b	19.9%	77.3%	2.8%		15.4%	48.5%	36.1%
LFS2002a	8.0%	91.0%	1.1%		11.6%	47.5%	40.9%
LFS2002b	16.0%	82.5%	1.5%		11.0%	49.6%	39.5%
LFS2003a	7.8%	90.4%	1.8%		12.2%	45.3%	42.5%
LFS2003b	9.9%	86.4%	3.8%		10.9%	45.5%	43.6%
LFS2004a	7.0%	88.4%	4.6%		10.4%	45.7%	43.9%
LFS2004b	14.3%	83.9%	1.9%		11.4%	44.3%	44.3%
LFS2005a	14.0%	83.2%	2.8%		11.6%	50.6%	37.8%
LFS2005b	17.5%	77.9%	4.7%		13.9%	47.0%	39.2%
LFS2006a	6.8%	91.5%	1.7%		14.8%	46.4%	38.8%
LFS2006b	12.0%	86.4%	1.6%		18.9%	41.2%	39.9%

3.4.2 Declaration of actual earnings amount versus earnings category

All the surveys under investigation in this paper – with the exception of OHS1996 – provide respondents with the choice of reporting their income either as an actual value or within an income category. Income categories are brackets within which a respondent's income falls. Various reasons exist for respondents answering in bands, such as high income earners not wanting to declare the actual value of their earnings, or respondents' uncertainty about the income of other household members. (Von Fintel (2006, 2007) analyses the use of such income categories in the South African context.)

Table 26 reports the percentage of the employed who choose to report their earnings from their main job in the two different ways discussed above. It is obvious the employees are more likely to declare actual amount. Besides, with the exception of LFS2000b, the percentage of employees reporting an actual amount stabilizes at approximately 75% in all LFSs.

Table 26: The option chosen by the respondents to declare earnings from main job, 1995 – 2006

	Employee		Self-Employed		All	
	Actual amount	**Income category**	**Actual amount**	**Income category**	**Actual amount**	**Income category**
OHS1995	77%	23%	70%	30%	76%	24%
OHS1996	0%	100%	0%	100%	0%	100%
OHS1997	65%	35%	65%	35%	65%	35%
OHS1998	58%	42%	51%	49%	57%	43%
OHS1999	51%	49%	39%	61%	49%	51%
LFS2000a	75%	25%	38%	62%	65%	35%
LFS2000b	89%	11%	53%	47%	81%	19%
LFS2001a	78%	22%	59%	41%	73%	27%
LFS2001b	78%	22%	60%	40%	75%	25%
LFS2002a	75%	25%	44%	56%	69%	31%
LFS2002b	74%	26%	52%	48%	70%	30%
LFS2003a	74%	26%	54%	46%	70%	30%
LFS2003b	71%	29%	50%	50%	67%	33%
LFS2004a	71%	29%	54%	46%	68%	32%
LFS2004b	73%	27%	56%	44%	69%	31%
LFS2005a	73%	27%	50%	50%	69%	31%
LFS2005b	72%	28%	61%	39%	70%	30%
LFS2006a	74%	26%	52%	48%	69%	31%
LFS2006b	75%	25%	58%	42%	71%	29%

3.4.3 Real gross monthly earnings from the main job[21]

Inconsistencies in wage data may have a significant impact on statistical and economic analysis done using the data because "if the analysis of wage trends is sensitive to presence of certain observations and some these are not representative of an underlying data generating process, then including the latter risks misleading conclusions" (Burger & Yu, 2006: 2). Burger and Yu (2006) also provide a detailed explanation of the impact of the outliers on the average real month earnings of the employed from the main job. For the remainder of Section 3, monthly real earnings of more than R500,000 will be regarded as outliers and excluded from tabulations.

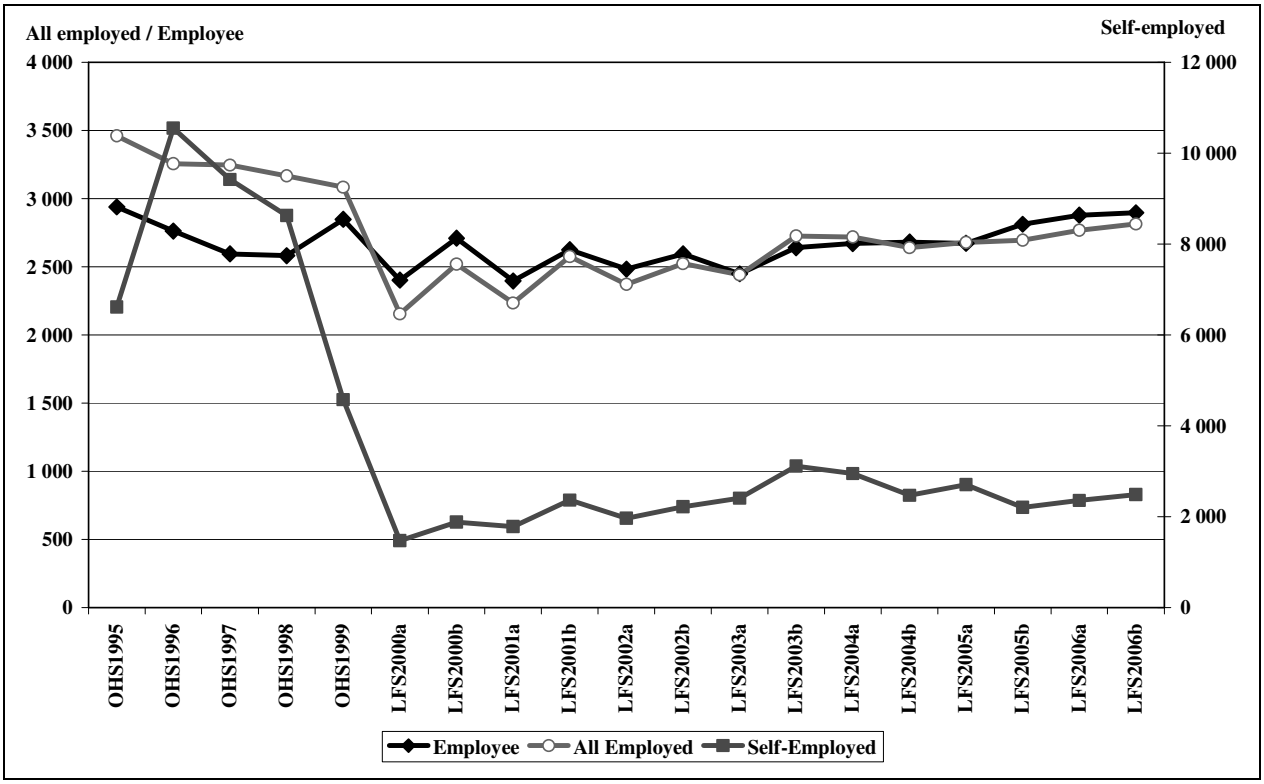
Table 27 presents the mean and standard deviation of the gross real monthly earnings of the employed. The self-employed appear to have unusually high mean incomes (as well as large standard deviations) in all OHSs. This contributes to the higher mean income of the employed in those years, which is illustrated in Figure 5. However, as pointed out earlier in Section 3.2.1, the self-employed account for a much smaller proportion of the employed in the OHS years (less than 15%).

[21] Nominal earnings were converted into real earnings (expressed in 2000 prices) using the South African Reserve Bank's CPI series (KBP7032N).

Table 27: Mean and standard deviation of the gross real monthly earnings of employed (Rand), 1995 – 2006

	Employee		Self-Employed		All	
	Mean	**Std Dev.**	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**
OHS1995	2,939	3,561	6,612	20,730	3,461	8,577
OHS1996	2,762	3,497	10,550	25,760	3,256	7,560
OHS1997	2,595	3,273	9,417	23,962	3,245	8,268
OHS1998	2,580	5,110	8,626	20,812	3,167	8,294
OHS1999	2,848	9,038	4,574	12,251	3,083	9,559
LFS2000a	2,401	5,325	1,471	7,285	2,153	5,925
LFS2000b	2,710	6,095	1,877	5,560	2,519	5,986
LFS2001a	2,396	3,191	1,780	6,258	2,234	4,230
LFS2001b	2,624	5,378	2,362	5,293	2,575	5,363
LFS2002a	2,483	3,530	1,966	5,508	2,371	4,047
LFS2002b	2,595	5,959	2,218	6,124	2,523	5,993
LFS2003a	2,449	3,511	2,405	6,913	2,441	4,339
LFS2003b	2,640	3,938	3,113	15,100	2,726	7,345
LFS2004a	2,670	3,731	2,950	6,699	2,718	4,394
LFS2004b	2,680	3,858	2,470	5,458	2,641	4,208
LFS2005a	2,672	3,831	2,702	6,567	2,679	4,501
LFS2005b	2,812	4,246	2,207	5,782	2,695	4,595
LFS2006a	2,878	4,107	2,355	6,011	2,768	4,580
LFS2006b	2,896	4,236	2,484	6,686	2,814	4,827

Figure 5: Mean monthly gross real earnings from main job by employment type (Rand), 1995 – 2006



3.3.4 Non-zero earnings of the unemployed and economically inactive people

Table 28 below reports the number of unemployed or economically inactive people in each sample who report non-zero earnings. From the table, it may be seen that a greater number of unemployed or economically inactive report non-zero earnings in OHS1995 and OHS1996.

Table 28: Number of unemployed/economically inactive people in the sample with non-zero earnings (unweighted), 1995 – 2006

	A	B	A/B
OHS1995	1,107	49,852	2.2%
OHS1996	963	29,443	3.3%
OHS1997	42	55,871	0.1%
OHS1998	31	32,331	0.1%
OHS1999	94	40,278	0.2%
LFS2000a	36	13,074	0.3%
LFS2000b	7	37,704	0.0%
LFS2001a	14	38,475	0.0%
LFS2001b	8	40,495	0.0%
LFS2002a	3	41,009	0.0%
LFS2002b	4	39,216	0.0%
LFS2003a	5	39,124	0.0%
LFS2003b	1	38,413	0.0%
LFS2004a	12	38,326	0.0%
LFS2004b	6	43,350	0.0%
LFS2005a	4	43,055	0.0%
LFS2005b	7	42,182	0.0%
LFS2006a	6	41,466	0.0%
LFS2006b	2	40,477	0.0%

A: Unemployed / Inactive with non-zero earnings
B: Total number of unemployed / inactive in the sample

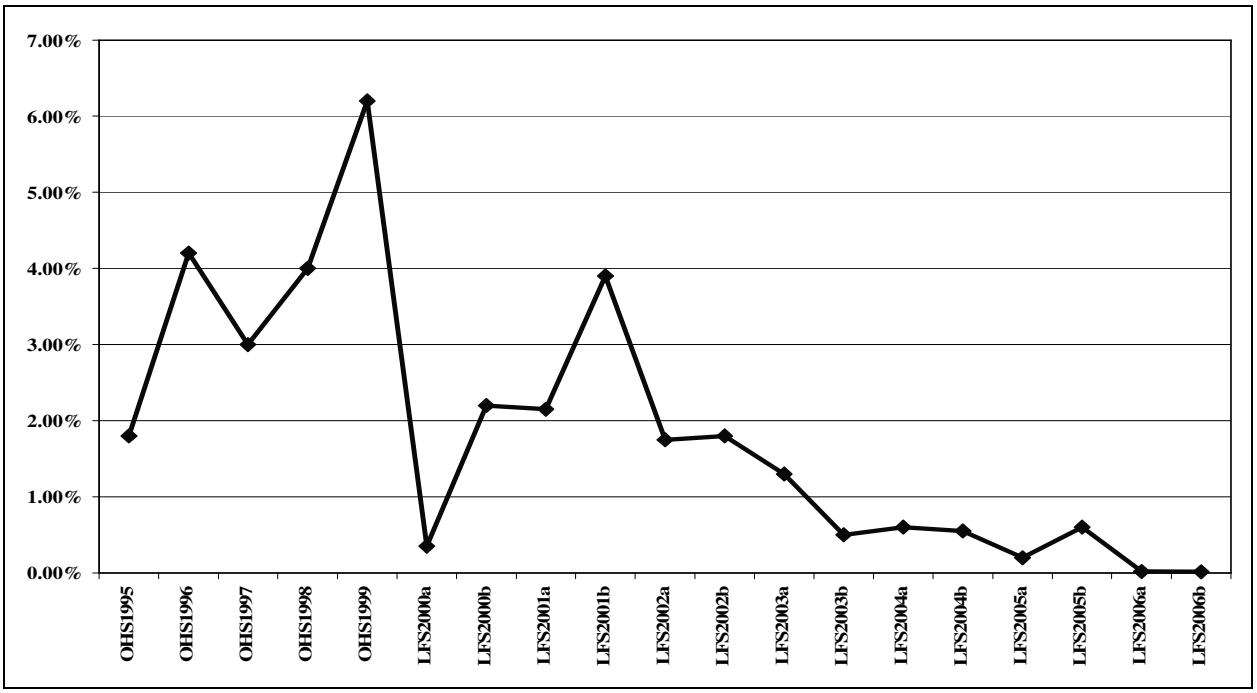
4. INCONSISTENCIES RESULTING FROM QUESTION FORMULATIONS

In this section, all question numbers refer the questionnaire for LFS2006b.

4.1 ACTION TO SEEK WORK

Question 3.2 reads "Why did …… not work during the past seven days?" It is suggested that if the respondent answers that he/she did not work in the past seven days because he/she is a student, housewife or retired, then the respondent should not be allowed to continue to respond to questions from question 3.8 until the end of section 3, since these questions pertain to action to seek work. It is obvious from their answer to question 3.2 that they "prefer" not to work or to seek work. However, the word "prefer" could be confusing or misleading to respondents. Inconsistencies occur because respondents who answer question 3.2 by saying that they prefer not to work or to seek work later report that they have taken action to seek work in question 3.8. Figure 6 provides a graphical example of such an inconsistency, showing the proportion of retired respondents giving inconsistent answers with regards to work-seeking action. This contradiction seems to less happen since LFS2003b, and has seemingly disappeared in 2006.

Figure 6: Proportion of retired giving contradictory answers regarding work-seeking action, 1995 – 2006



4.2 FIRM SIZE

The question regarding firm size in LFS2005b (question 4.16) reads "How many regular workers has the organization/business/enterprise/branch where . . . . works, including him/herself?" The question does not however define the meaning of "worker" clearly. That is, the question does not define whether the employer is classified as a worker. Table 29 reports employees' answers on firm size. It seems peculiar that roughly 10% of employees report that here is only one worker in the organization, since if the question assumes that the respondent will count the employer as a worker, there must be at least two workers (1 employer and 1 employee) in the organization.

Table 29: Employees' answers on firm size, 2000 – 2006

	1 person	2-4 persons	5-9 persons	10-19 persons	20-49 persons	50 or more
LFS2000a	13%	12%	11%	13%	15%	35%
LFS2000b	12%	12%	12%	15%	16%	32%
LFS2001a	11%	12%	11%	14%	17%	34%
LFS2001b	12%	11%	11%	15%	17%	35%
LFS2002a	12%	11%	11%	15%	17%	33%
LFS2002b	11%	10%	10%	14%	18%	36%
LFS2003a	12%	10%	10%	15%	19%	35%
LFS2003b	12%	10%	11%	15%	19%	33%
LFS2004a	11%	9%	10%	15%	20%	35%
LFS2004b	12%	7%	7%	8%	24%	42%
LFS2005a	11%	10%	10%	15%	20%	34%
LFS2005b	10%	11%	11%	16%	19%	33%
LFS2006a	11%	10%	11%	15%	20%	33%
LFS2006b	11%	10%	10%	15%	19%	36%

5. COMMENTS ON QUESTIONNAIRE AND RECOMMENDATIONS FOR IMPROVEMENT

In this section, all question numbers refer the questionnaire for LFS2006b, unless stated otherwise.

5.1 SUGGESTIONS FOR SECTION 1

5.1.1 Education

It is suggested that question 1.8 ("Which of the following educational institutions, if any, does . . . . currently attend?") should be asked directly after question 1.3 (1.3a "What is the highest level of education that . . . . . has successfully completed?" and 1.3b "In what area was the highest diploma, certificate or degree?"). This may serve to lessen confusion and inconsistency. It seems slightly less logical to place the questions pertaining to training and language (questions 1.4 to 1.6) between the two aforementioned questions on educational attainment.

5.1.2 Language

It will be more logical if questions 1.7a and 1.7b ("Can . . . . . read in at least one language?" and "Can . . . . . write in at least one language?" respectively) are asked directly after question 1.2 ("Which language does . . . . speak most often at home?").

5.1.3 Training

The meaning of "training" is unclear in questions 1.4 and 1.5 (which are respectively "Has . . . . been trained in skills that can be used for work, e.g. book-keeping, security guard training, welding, child-minding?" and "The last time . . . . received this type of training, how long did it last?"). It is not clear whether this training refers to an academic course being completed by the respondents to gain the skills, or whether it refers to on-the-job training of workshop training by the employer. It is therefore suggested that the meaning of training be clarified.

5.1.4 Breadwinner

Question 1.15 ("Who is the person that usually brings the most money in to the household?") should perhaps be asked in the household section of the questionnaire.

5.2 SUGGESTIONS ON SECTION 3

Section 3 in the questionnaire is the section that is answered by respondents who did not work and who were not absent from work in the last 7 days.

5.2.1 Odd jobs

The options from which respondents must choose to answer question 3.1 ("How does . . . . . support him/herself if not working for at least 1 hour in the last 7 days?") appear to be ambiguous and may therefore prove confusing for respondents. As mentioned before, it is clearly stated at the beginning of section 3 that this section may only be answered by respondents who did not

work and who were not absent from work in the 7 days prior to answering the questionnaire. One of the options for question 3.1 is "did odd jobs" which implies that the respondents was in fact working in the last 7 days and therefore, strictly speaking, the respondent should not be allowed to take part in section 3.

Stats SA takes this into consideration by asking the respondent to return to section 2 if his/her answer to question 3.1 is "did odd jobs". If this is the case, the respondent will be classified as being employed by StatsSA. However, it is suggested that "odd jobs" should be added as an option to question 2.1 ("In the last 7 days, did . . . . . do any of the following activities, even for only one hour?"). Further, "odd jobs" should be clearly defined. The options for question 2.1 should therefore read

a) Run or do any kind of business, big or small, for himself/herself or with one or more partners?
……..
h) Do any odd jobs (which are defined as . . . . . .)
i) Beg for money or food in public?

It is important to point out however that in almost all the years under investigation, fewer than 100 respondents reported "did odd jobs" when answering question 3.1.

5.2.2 How the unemployed support themselves

As mentioned before, it is possible that the respondent has more than one way of supporting himself/herself (question 3.1 gives the respondent numerous options from which to choose). It is therefore suggested than question 3.1 be rephrased so as to read "What major activity does . . . . . engage in to support himself/herself, if not working for at least 1 hour in the last 7 days?" or "Which of the following activities does . . . . . engage in to support himself/herself, if not working for at least 1 hour in the last 7 days? Rank these activities in descending order of importance with regards to the contribution these activities make to his/her financial support."

5.2.3 Action to look for work

It is surprising that 'looking for working in the newspaper', one of the most common ways of finding work, is not included as an option in question 3.9 ("In the past 4 weeks, what has . . . . . done to look for work or to start a business?"). Furthermore, technological advances imply that job searching may take place on the internet. It is therefore suggested that the options for question 3.9 might need to be reconsidered.

5.2.4 Duration of job search

It is suggested that options 7 and 8 for question 3.10 ("How long has . . . . . been trying to find work or start a business?") be broken down in to smaller categories. Options 7 and 8 are "1 year to less than 3 years" and "3 years or more" respectively. This suggestion is made on the grounds that close to two-thirds of respondents (excluding those who choose option 9 – "Don't know") report having been searching for work for at least 1 year, as shown in Table 30. Thus more cateogries at the longer end may give a better representation of the length of job search.

Table 30: Broadly unemployed's answers on duration of looking for work, excluding 'Don't know', 1995 – 2006

	< 1month	1 month to less than 6 months	6 months to less than 1 year	1 year to less than 3 years	3 years or more	Total		1 year or more
OHS1995	6%	11%	18%	29%	36%	100%		65%
OHS1996	9%	9%	16%	29%	37%	100%		66%
OHS1997	7%	11%	18%	28%	36%	100%		64%
OHS1998	9%	10%	17%	26%	37%	100%		63%
OHS1999	8%	9%	14%	27%	42%	100%		69%
LFS2000a	7%	15%	11%	31%	36%	100%		66%
LFS2000b	6%	11%	17%	30%	37%	100%		67%
LFS2001a	7%	15%	14%	27%	37%	100%		64%
LFS2001b	6%	16%	12%	28%	38%	100%		66%
LFS2002a	8%	18%	9%	27%	38%	100%		65%
LFS2002b	7%	15%	12%	26%	40%	100%		66%
LFS2003a	5%	17%	9%	26%	42%	100%		68%
LFS2003b	7%	17%	12%	26%	37%	100%		63%
LFS2004a	8%	17%	9%	26%	40%	100%		66%
LFS2004b	8%	17%	12%	24%	39%	100%		63%
LFS2005a	7%	19%	10%	25%	39%	100%		64%
LFS2005b	7%	18%	13%	25%	36%	100%		61%
LFS2006a	9%	21%	10%	25%	36%	100%		61%
LFS2006b	9%	21%	12%	23%	35%	100%		58%

5.2.5 Reasons for not trying to find work

Option 6 ("No job available in the area") of question 3.11 ("What was the main reason why . . . . . did not try to find work or start a business in the past four weeks?") is ambiguous. In order for an individual to discover that there are no jobs available within a given area, he must first engage in some kind action to ascertain whether there are jobs in the area. It is therefore suggested that option 6 is either dropped as a possible response to question 3.11 or re-phrased. Approximately 50% of respondents in all years reported that they had not taken action to seek work of start a business because there was no work in the area which may indicate a degree of confusion. It would be useful to know how respondents became aware of the lack of work in the given area.

5.2.6 Whether respondent ever worked before

Question 3.12 ("Has . . . . . ever worked for pay, profit or family gain?") should be re-phrased to take time in to consideration. For example, the question should read "Has . . . . . ever worked for pay, profit or family gain, even if only for 1 hour, and regardless of when it happened?"

5.2.7 How long since last worked

It is suggested that option 10 ("3 years or more") of question 3.13 ("How long ago was it since . . . . . last worked?") be broken down further, because roughly 50% of respondents in all years (excluding those who chose "Don't know") reported having last worked 3 years ago or longer. These results are reported in Table 31. Note that this question was only asked since OHS1999.

Table 31: Broadly unemployed's answers on when they last worked, excluding 'Don't know', 1999 – 2006

	More than 1 week to less than 1 month	1 month to less than 6 months	6 months to less than 1 year	1 year to less than 2 years	2 years to less than 3 years	3 years or more
OHS1999	4%	11%	10%	13%	11%	51%
LFS2000a	2%	11%	9%	13%	13%	52%
LFS2000b	3%	9%	9%	16%	14%	50%
LFS2001a	2%	14%	8%	12%	13%	51%
LFS2001b	2%	11%	8%	13%	12%	54%
LFS2002a	2%	13%	7%	12%	12%	54%
LFS2002b	2%	11%	9%	11%	11%	57%
LFS2003a	2%	13%	6%	12%	12%	56%
LFS2003b	2%	12%	8%	12%	11%	54%
LFS2004a	3%	13%	7%	11%	11%	56%
LFS2004b	2%	13%	8%	11%	11%	54%
LFS2005a	2%	14%	7%	10%	12%	55%
LFS2005b	3%	14%	8%	13%	10%	52%
LFS2006a	3%	14%	8%	11%	12%	52%
LFS2006b	3%	16%	9%	11%	10%	50%

5.2.8 Other suggestions

It is suggested that question 3.13 ("How long ago was it since . . . . . last worked?") be asked before question 3.10 ("How long has . . . . . been trying to find work or start a business?"). Perhaps a further improvement might be to ask question 3.13 before question 3.8 ("During the past four weeks, has . . . . . taken any action a) to look for any kind of work? b) to start any kind of business?").

5.3 SUGGESTIONS ON SECTION 4

5.3.1 Tenure

It seems strange that only respondents choosing option 1 or 2 ("Working for someone else for pay" and "Working for one or more private households as a domestic employee, gardener or security guard" respectively) (i.e. employees) to question 4.3 ("In . . . . . 's main job, was he/she …..") are allowed asked to move to question 4.5 ("When did . . . . . start working with the (main) employer mentioned above (firm, institution or private household)?"), but that respondents choosing options 3, 4 or 5 ("Working on his/her own or on a small household farm/plot or collecting natural products from the forest or seas", "Working on his/her own or with partner, in any type of business (including commercial farms)" and "Helping without pay in a household business" respectively) (i.e. self-employed) are not allowed to do so. It is suggested therefore that question 4.5 be rephrased along the lines of "If . . . . . is an employee, when did he/she start working with the (main) employer mentioned above (firm, institution or private household)? If . . . . . is self-employed, when did he/she start his/her current business?"

5.3.2 Employment type

Options 7 and 8 ("A private business or private household" and "Self-employed") to question 4.14 ("Is the business or enterprise/branch where . . . . . works ……..?") are confusing because if the business of a self-employed individual taken place in a private household (for example a car repair service run from the self-employed individual's home), then this individual falls into category 7 and 8 simultaneously.

5.4 SUGGESTIONS ON SECTION 7

LFS2004b is referred to when a question is mentioned here since the household section is not available since LFS2005b. LFS2005a does contain a household section but many questions contained in LFS2004b were no longer asked[22]. Section 7 covers the information about the household, such as dwelling type, and access to water and electricity.

5.4.1 Main source of household income

It is suggested that the options for question 7.29 ("What is the main source of income for this household?") include options like "Income from fixed deposits/savings", "Income from property/investment", etc. Further, this question does not allow for the income from property, such as rent received from letting a house. It is suggested that income from property be added as an option for question 7.29.

[22] In OHS1993, section 1 asks a few questions about the household, including dwelling type, number of rooms, sanitation, water and energy sources. In OHS1994 and OHS1995, there is a section called 'information regarding dwelling and perceived quality of life', which asks relatively more questions about the household. In OHS1996 – OHS1999, LFS2000b – LFS2001b, LFS2002b, LFS2003b and LFS2004b, there is a section which is clearly named as 'information regarding the household', which asks a lot of questions, including dwelling type, ownership of dwelling, number of room, main source of water, energy, sanitation and refuse removal, main source of income, access to grants, monthly household expenditure, etc. Although LFS2004a and LFS2005a contain a household section, a lot of questions are not asked. Finally, there is no household section in LFS2000a, LFS2002a, LFS2003a, and LFS2005b – LFS2006b.

5.4.2 Household expenditure categories

Table 32 below presents household expenditure categories in selected LFSs[23]. Almost 70% of all households fall in the first three groups (i.e. R0 – R399, R400 – R799, R800 – R1199), suggesting that the expenditure categories are rather uninformative in their present state. It is therefore apparent that categories that allow for finer distinctions are required, such as is the case for earnings.

Table 32: Household monthly expenditure, excluding 'Don't know' and 'Refuse', selected LFSs

	Black				**White**				**All households**			
	LFS 2001b	LFS 2002b	LFS 2003b	LFS 2004b	LFS 2001b	LFS 2002b	LFS 2003b	LFS 2004b	LFS 2001b	LFS 2002b	LFS 2003b	LFS 2004b
R0 – R399	36%	33%	27%	23%	1%	1%	0%	0%	29%	27%	22%	19%
R400 – R799	33%	33%	36%	36%	3%	2%	1%	2%	28%	29%	30%	30%
R800 – R1199	14%	15%	18%	18%	6%	4%	4%	4%	13%	14%	16%	16%
R1200 – R1799	7%	8%	8%	9%	8%	8%	4%	4%	8%	8%	9%	9%
R1800 – R2499	4%	5%	5%	6%	11%	12%	9%	9%	6%	6%	6%	7%
R2500 – R4999	4%	4%	5%	5%	32%	26%	24%	23%	9%	8%	9%	9%
R5000 – R9999	1%	1%	2%	2%	26%	33%	34%	34%	5%	5%	6%	7%
R10000+	0%	1%	0%	1%	13%	15%	23%	24%	2%	2%	3%	3%
% of households in the first three expenditure categories	83%	81%	81%	77%	10%	7%	5%	6%	70%	70%	68%	65%

5.4.3 Other suggestions

It would be useful to include the question about the relationship of the respondent to the household head. This question was asked in all the OHSs, but not anymore since LFS2000a.

5.5 HOUSEHOLD WEIGHTS

Table 33 presents the number of households and the racial share of households from OHS1997 to LFS2005a. Note that these are the only surveys that contain the household weight variable in the datasets. From the table, it is clear that the household weight in OHS1999 is inaccurate, since OHS1999 undoubtedly under-estimates the number of households relative to all the other years reported and is inaccurate with the estimation of the racial shares of the households.

[23] The question on monthly household expenditure was only asked in OHS1999, LFS2001b, LFS2002b, LFS2003b and LFS2004b. Unfortunately, the household weight in OHS1999 is unreliable, to be explained in section 5.5.

Table 33: Number of households and racial share (weighted), 1997 – 2005

	Number of households (weight = household weight)					
	Black	**Coloured**	**Indian**	**White**	**Others**	**Total**
OHS1997	6,734,717	759,434	247,933	1,514,623	0	9,256,707
OHS1998	6,739,183	758,780	247,932	1,515,871	14,355	9,276,121
OHS1999	3,154,869	644,535	237,565	2,099,579	11,977	6,148,525
LFS2000b	8,518,652	915,104	283,891	1,610,449	24,069	11,352,165
LFS2001a	8,666,875	933,376	295,883	1,617,219	13,792	11,527,145
LFS2001b	8,660,806	900,884	295,041	1,555,636	14,520	11,426,887
LFS2002b	8,896,382	909,591	296,072	1,550,053	11,794	11,663,892
LFS2003b	9,438,025	954,988	317,714	1,593,302	6,899	12,310,928
LFS2004a	9,422,325	968,340	320,487	1,568,406	7,562	12,287,120
LFS2004b	9,640,812	952,323	320,558	1,632,703	27,347	12,573,743
LFS2005a	9,652,936	1,000,767	323,492	1,588,681	16,227	12,582,103
Racial distribution						
	Black	**Coloured**	**Indian**	**White**	**Others**	**Total**
OHS1997	72.8%	8.2%	2.7%	16.4%	0.0%	100.0%
OHS1998	72.7%	8.2%	2.7%	16.3%	0.2%	100.0%
OHS1999	51.3%	10.5%	3.9%	34.1%	0.2%	100.0%
LFS2000b	75.0%	8.1%	2.5%	14.2%	0.2%	100.0%
LFS2001a	75.2%	8.1%	2.6%	14.0%	0.1%	100.0%
LFS2001b	75.8%	7.9%	2.6%	13.6%	0.1%	100.0%
LFS2002b	76.3%	7.8%	2.5%	13.3%	0.1%	100.0%
LFS2003b	76.7%	7.8%	2.6%	12.9%	0.1%	100.0%
LFS2004a	76.7%	7.9%	2.6%	12.8%	0.1%	100.0%
LFS2004b	76.7%	7.6%	2.5%	13.0%	0.2%	100.0%
LFS2005a	76.7%	8.0%	2.6%	12.6%	0.1%	100.0%

6. CONCLUSION

The paper has discussed the sample size of the OHSs and LFSs (focusing on the working age population), the inconsistencies that occur in the data independently of the way in which the questions were asked (focusing on the general, employment and earnings variables), the data inconsistencies that resulted from the way in which the questions were formulated or placed in a given sequence, and the suggestions on adjustments that should be made to improve the consistency in the responses.

The paper has shown that these inconsistencies have become less of a problem with more recent datasets, although there have been some recent changes to the survey questionnaire that have made comparison across surveys somewhat more difficult. Generally, however, the quality of data and comparability across time are improving, thus allowing researchers and policy makers to obtain increasingly reliable and credible labour market trends from these datasets.

7. Bibliography

Burger, R.P. & Yu, D. 2006. *Wage Trends in Post-Apartheid South Africa: Constructing an Earnings Series from Household Survey Data*. Stellenbosch University, Department of Economics, Working Paper 04/2006.

Statistics South Africa, 1999. *October Household Survey 1996 Statistical release*. Government printer, Pretoria.

Von Fintel, Dieter. 2006. *Earnings bracket obstacles in household surveys – How sharp are the tools in the shed?* Stellenbosch University, Department of Economics, Working Paper 08/2006.

Von Fintel, Dieter. 2007. Dealing with earnings bracket responses in household surveys - How sharp are midpoint imputations? *South African Journal of Economics*. Vol 75 No.2: 293-312.

APPENDIX

Table A.1: Percentage of people in each age group, 1993 – 2006

	0-4 yrs	5-9 yrs	10-14 yrs	15-19 yrs	20-24 yrs	25-29 yrs	30-34 yrs	35-39 yrs	40-44 yrs	45-49 yrs	50-54 yrs	55-59 yrs	60-64 yrs	65+ yrs	Unspecified	Population
OHS1993	12.6%	11.8%	10.9%	10.4%	9.6%	8.5%	7.5%	6.3%	5.2%	4.3%	3.6%	2.9%	2.2%	4.2%	0.0%	32,207,758
OHS1994	12.8%	12.1%	11.4%	10.4%	9.4%	8.2%	7.2%	6.3%	5.2%	4.2%	3.4%	2.8%	2.2%	4.4%	0.0%	40,251,142
OHS1995	10.2%	12.4%	12.1%	10.4%	10.2%	8.5%	7.7%	6.5%	5.1%	4.0%	3.0%	2.7%	2.2%	4.9%	0.0%	39,659,831
OHS1996	11.0%	11.3%	11.9%	10.4%	10.4%	8.2%	7.7%	6.6%	5.2%	4.1%	3.1%	2.7%	2.4%	4.9%	0.0%	40,582,538
OHS1997	11.5%	11.3%	11.4%	10.4%	9.8%	8.7%	7.7%	6.6%	5.4%	4.3%	3.2%	2.7%	2.2%	4.8%	0.0%	41,443,101
OHS1998	10.2%	12.4%	12.1%	10.4%	10.2%	8.5%	7.6%	6.5%	5.1%	4.0%	3.0%	2.7%	2.2%	4.9%	0.0%	42,235,733
OHS1999	10.2%	12.4%	12.1%	10.4%	10.2%	8.5%	7.6%	6.5%	5.1%	4.0%	3.0%	2.7%	2.2%	4.9%	0.2%	43,271,686
LFS2000a	10.2%	12.4%	12.1%	10.4%	10.1%	8.5%	7.6%	6.5%	5.2%	4.0%	3.0%	2.7%	2.2%	4.9%	0.2%	43,620,361
LFS2000b	11.3%	11.4%	11.0%	10.6%	10.0%	9.3%	7.0%	6.1%	5.4%	4.4%	3.6%	2.8%	2.5%	4.4%	0.2%	44,821,345
LFS2001a	11.2%	11.4%	11.0%	10.5%	10.0%	9.3%	7.1%	6.0%	5.4%	4.4%	3.6%	2.9%	2.5%	4.5%	0.2%	45,080,410
LFS2001b	11.2%	11.4%	11.0%	10.5%	10.0%	9.3%	7.1%	6.1%	5.4%	4.4%	3.6%	2.9%	2.5%	4.5%	0.2%	45,081,045
LFS2002a	11.1%	11.3%	11.0%	10.5%	10.0%	9.3%	7.3%	6.0%	5.4%	4.5%	3.6%	2.9%	2.5%	4.5%	0.1%	45,324,735
LFS2002b	11.1%	11.2%	11.0%	10.4%	10.0%	9.2%	7.4%	6.0%	5.4%	4.5%	3.6%	2.9%	2.5%	4.6%	0.1%	45,560,990
LFS2003a	11.0%	11.2%	11.0%	10.4%	10.0%	9.2%	7.6%	5.9%	5.4%	4.6%	3.7%	2.9%	2.5%	4.7%	0.1%	45,810,074
LFS2003b	10.9%	11.1%	10.9%	10.4%	10.0%	9.1%	7.7%	5.9%	5.4%	4.6%	3.7%	3.0%	2.6%	4.7%	0.1%	46,046,026
LFS2004a	10.9%	11.0%	10.9%	10.4%	10.0%	9.1%	7.8%	5.9%	5.4%	4.6%	3.7%	3.0%	2.6%	4.8%	0.1%	46,270,894
LFS2004b	10.8%	10.9%	10.9%	10.4%	9.9%	9.0%	7.9%	5.9%	5.3%	4.6%	3.7%	3.0%	2.6%	4.9%	0.1%	46,490,122
LFS2005a	10.8%	10.8%	10.9%	10.4%	9.9%	9.0%	8.0%	5.9%	5.3%	4.7%	3.7%	3.0%	2.6%	4.9%	0.1%	46,699,967
LFS2005b	10.8%	10.7%	10.8%	10.4%	9.8%	9.0%	8.0%	5.9%	5.3%	4.7%	3.8%	3.0%	2.7%	5.0%	0.2%	46,917,195
LFS2006a	10.9%	10.6%	10.8%	10.4%	9.8%	9.0%	8.1%	5.9%	5.1%	4.7%	3.8%	3.1%	2.7%	5.0%	0.1%	47,184,311
LFS2006b	10.9%	10.5%	10.7%	10.4%	9.8%	9.0%	8.1%	6.0%	5.1%	4.7%	3.9%	3.1%	2.7%	5.1%	0.1%	47,429,106

Table A.2: Number of individuals, by marital status (after correcting the OHS1998 mistakes)

	Never married	Married – civil	Married – traditional	Married	Live together	Married or live together	Widow/ Widower	Divorced	Unspecified	Total
OHS1993	9,341,338			8,327,533	779,098	9,106,631	690,488	488,240	0	19,626,697
OHS1994	11,958,330	7,322,126	2,439,359		895,926	10,657,411	833,792	625,035	0	24,074,568
OHS1995	12,243,524	6,770,378	2,778,450		947,571	10,496,399	854,365	596,295	0	24,190,583
OHS1996	12,535,479	6,811,134	2,912,855		1,032,953	10,756,942	900,652	613,364	102,628	24,909,065
OHS1997	13,261,999	6,927,977	2,468,406		1,239,226	10,635,609	918,992	689,489	0	25,506,089
OHS1998	13,146,937	6,427,808	2,777,365		1,606,440	10,811,613	913,351	724,775	68,557	25,665,233
OHS1999	13,799,111	6,456,863	2,654,271		1,519,089	10,630,223	856,046	794,043	167,122	26,246,545
LFS2000a	14,226,190					10,321,789	986,732	913,407	16,992	26,465,110
LFS2000b	14,488,583					11,294,393	1,166,684	848,118	38,678	27,836,456
LFS2001a	14,596,953					11,433,827	1,180,631	844,294	6,299	28,062,004
LFS2001b	14,761,635					11,349,101	1,156,564	813,558	3,469	28,084,327
LFS2002a	14,766,533					11,539,513	1,142,799	845,979	3,431	28,298,255
LFS2002b	14,884,904					11,619,814	1,121,950	864,934	3,486	28,495,088
LFS2003a	15,186,513					11,531,044	1,190,070	808,205	8,689	28,724,521
LFS2003b	15,223,081					11,608,662	1,212,947	856,306	5,234	28,906,230
LFS2004a	15,540,497					11,536,542	1,176,121	837,230	9,397	29,099,787
LFS2004b	15,317,504			9,318,125	2,573,501	11,891,626	1,199,112	860,929	1,650	29,270,821
LFS2005a	15,750,154			9,188,507	2,583,067	11,771,574	1,148,727	808,152	11,156	29,489,763
LFS2005b	15,985,864			8,887,407	2,802,347	11,689,754	1,183,837	799,814	4,110	29,663,379
LFS2006a	16,417,448			8,936,392	2,513,902	11,450,294	1,185,573	764,198	311	29,817,824
LFS2006b	16,398,135			8,861,800	2,712,459	11,574,259	1,265,965	732,812	1,400	29,972,521

Table A.3: Educational attainment question: OHS1993 – OHS1994

Category	eduyear
0: No schooling	0
1: Sub A/Sub B/Grade 1/Grade 2/Std 1	2
2: Std 2	4
3: Std 3	5
4: Std 4	6
5: Std 5	7
6: Std 6	8
7: Std 7	9
8: Std 8	10
9: Std 9	11
10: Std 10	12
11: Diploma/certificate with Std 9 or lower	11
12: Diploma/certificate with Std 10	13
13: Degree/equivalent to a 3 year academic training after Std 10	15
14: Other	???
15: Unspecified	???

Table A.4: Educational attainment question: OHS1995

Category	eduyear
00: No schooling	0
01: Sub A/Sub B/Grade 1/Grade 2/Std 1	2
02: Std 2	4
03: Std 3	5
04: Std 4	6
05: Std 5	7
06: Std 6	8
07: Std 7	9
08: Std 8/NTC I	10
09: Std 9/NTC II	11
10: Std 10/NTC III	12
11: Diploma/certificate with Std 9 or lower	11
12: Diploma/certificate with Std 10	13
13: Degree	15
14: Other	???
15: Unspecified	???

Table A.5: Educational attainment question: OHS1996

Category	eduyear
00: No schooling	0
01: Sub A/Grade 1	1
02: Sub B/Grade 2	2
03: Std 1	3
04: Std 2	4
05: Std 3	5
06: Std 4	6
07: Std 5	7
08: Std 6/Form 1	8
09: Std 7/Form 2	9
10: Std 8/Form 3	10
11: Std 9/Form 4	11
12: Std 10/Form 5	12
13: NTC I	10
14: NTC II	11
15: NTC III	12
16: Diploma/certificate with Std 9 or lower	11
17: Diploma/certificate with Std 10	13
18: Degree	15
19: Other	???

Table A.6: Educational attainment question: OHS1997 – OHS1998

QUESTION 1: HIGHEST SCHOOL STANDARD PASSED	
Category	**Years at school**
00: None	0
01: Grade 0	0
02: Grade 1/Sub A	1
03: Grade 2/Sub B	2
04: Grade 3/Std 1/Level 1	3
05: Grade 4/Std 2/Level 2	4
06: Grade 5/Std 3/Level 3	5
07: Grade 6/Std 4/Level 4	6
08: Grade 7/Std 5/Level 5	7
09: Grade 8/Std 6/Level 6/Form I	8
10: Grade 9/Std 7/Level 7/Form II	9
11: Grade10/Std 8/Level 8/Form III/NTC 1/RCE Higher	10
12: Grade 11/City of Guilds Inter grade/O Levels/College of Perceptions/ COP/CDE/NTC 2/ Std 9/GCE/General Certificate of Education/Certificate of Secondary Education/ Form IV	11
13: City of Guilds Final/O, M and A Levels/S Levels/M Levels/A Levels/ O and M Levels/O and A Levels/NTC 3/Grade 12/Std 10/Senior Certificate/Matric/Law Matric/Abitur/Subsidiary/Form V	12
99: Unspecified / not reported	???
QUESTION 2: HIGHEST TERTIARY QUALIFICATION	
Category	**eduyear**
1: Certificate	Years at school + 1 (Exception: when years of schooling = 11, then eduyear remains at 11)
2: Diploma only	Years at school + 1 (Exception: when years of schooling = 11, then eduyear remains at 11)
3: Bachelor's	15 (regardless of what happens to years at school)*
4: Bachelor's + Diploma	16 (regardless of what happens to years at school)*
5: Bachelor's + Honours	16 (regardless of what happens to years at school)*
6: Master's	17 (regardless of what happens to years at school)*
7: Doctor's	20 (regardless of what happens to years at school)*
8: Other	Years at school
9: Unspecified / not reported	???

* It is assumed that if the respondent declares he has obtained at least Bachelor Degree in the question on the highest tertiary qualification, but also claims he has never passed Matric in the question on the highest school standard passed, the answer of the former is more reliable.

Table A.7: Educational attainment question: OHS1999 – LFS2004a

Category	eduyear
00: No Schooling	0
01: Grade 0	0
02: Sub A/Grade 1	1
03: Sub B/Grade 2	2
04: Grade 3/Standard 1	3
05: Grade 4/Standard 2	4
06: Grade 5/Standard 3	5
07: Grade 6/Standard 4	6
08: Grade 7/Standard 5	7
09: Grade 8/Standard 6/Form 1	8
10: Grade 9/Standard 7/Form 2	9
11: Grade 10/Standard 8/Form 3	10
12: Grade 11/Standard 9/Form 4	11
13: Grade 12/Standard 10/Form 5/Matric	12
14: NTC I	10
15: NTC II	11
16: NTC III	12
17: Diploma/Certificate with less than Grade 12/Std 10	11
18: Diploma/Certificate with Grade 12/Std 10	13
19: Degree	15
20: Postgraduate Degree or Diploma	16
21: Other, Specify	???
22: Don't Know	???

Table A.8: Educational attainment question: LFS2004b – LFS2006b

Category	eduyear
00 = No Schooling	0
01 = Grade R/0	0
02 = Grade 1/Sub A	1
03 = Grade 2/Sub B	2
04 = Grade 3/Standard 1	3
05 = Grade 4/Standard 2	4
06 = Grade 5/Standard 3	5
07 = Grade 6/Standard 4	6
08 = Grade 7/Standard 5	7
09 = Grade 8/Standard 6/Form 1	8
10 = Grade 9/Standard 7/Form 2	9
11 = Grade 10/Standard 8/Form 3	10
12 = Grade 11/Standard 9/Form 4	11
13 = Grade 12/Standard 10/Form 5/Matric	12
14 = NTC I	10
15 = NTC II	11
16 = NTC III	12
17 = Certificate with Less than Grade 12/Std 10	11
18 = Diploma with Less than Grade 12/Std 10	11
19 = Certificate with Grade 12/Std 10	13
20 = Diploma with Grade 12/Std 10	13
21 = Bachelors Degree	15
22 = Bachelors Degree and Diploma	16
23 = Honours Degree	16
24 = Higher Degree (Masters, Doctorate)	17
25 = Other	???
26 = Don’t Know	???

Table A.9: Current education status question: OHS1993 – LFS2006b

OHS1993 Is … presently attending school/college/university, etc. on a full-time basis?* 1: Yes 2: No
OHS1994 – OHS1995 Is … presently attending school/college/university/technikon, etc.? 1: Yes, full-time 2: Yes, part-time 3: No
OHS1996 – OHS1998 Does (the person) presently attend school, college, technikon or university? (This includes study by correspondence but excludes crèche and pre-school) 1: Yes, full-time 2: Yes, part-time 3: No
OHS1999 Which of the following educational institutions, if any, does…… attend? (Include distance and correspondence education) (Could be either full-time or part-time) 1: School 2: University 3: Technikon 4: College 5: Adult basic education and training/literacy classes 6: Other adult education classes 7: Other than any of the above
LFS2000a – LFS2003b, LFS2004b Which of the following educational institutions, if any, does currently …… attend? (Include distance and correspondence education) (Could be either full-time or part-time) 1: School 2: University 3: Technikon 4: College 5: Adult basic education and training/literacy classes 6: Other adult education classes 7: Other than any of the above 8: None
LFS2004a, LFS2005a – LFS2006b Which of the following educational institutions, if any, does currently …… attend? (Include distance and correspondence education) (Could be either full-time or part-time) 1: Pre-school /crèche 2: School 3: University 4: Technikon 5: College 6: Adult basic education and training/literacy classes 7: Other adult education classes 8: Other than any of the above 9: None

* It seems the problem in OHS1993 is that it is impossible to capture people who were attending education on a part-time basis at the time of the survey.

Table A.10: The answer that must be provided by the respondents before they could be qualified as employed immediately, OHS1993 – LFS2006b

OHS1993 Now I am going to ask questions about … activities. What did … do most during the last 7 days? 1: Working
OHS1994 Now I am going to ask questions about … activities. What did … do most during the last 7 days? 1: Working
OHS1995 – OHS1996 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
OHS1997 – OHS1998 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
OHS1999 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
LFS2000a – LFS2006b 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?

Table A.11: Number of people who are qualified as employed immediately, OHS1995 – LFS2006b

	Full-time	Part-time	Casual/ Seasonal					Total	Total employed
OHS1995	8,672,177	827,170	N/A					9,499,347	9,499,347
OHS1996	7,924,672	959,447	N/A					8,884,119	8,966,307
OHS1997	8,036,972	631,196	341,645					9,009,813	9,093,647
OHS1998	8,179,686	693,336	437,567					9,310,589	9,370,130
OHS1999	8,449,934	1,044,780	792,647					10,287,361	10,356,143
	(a)	**(b)**	**(c)**	**(d)**	**(e)**	**(f)**	**(g)**	**Total: choosing at least one option**	**Total employed**
LFS2000a	1,646,332	8,117,565	1,088,499	220,044	1,978,098	217,903	57,018	11,748,664	11,874,409
LFS2000b	1,920,304	8,394,903	982,359	222,687	1,145,090	112,448	30,924	12,128,390	12,224,406
LFS2001a	2,634,597	7,989,226	986,054	229,823	788,802	39,912	15,648	12,047,487	12,260,207
LFS2001b	1,816,477	7,946,188	1,022,351	155,923	324,034	74,977	10,213	11,047,770	11,167,541
LFS2002a	1,670,042	7,968,650	1,054,235	113,889	879,397	48,955	14,118	11,456,782	11,603,398
LFS2002b	1,708,028	8,018,022	1,014,887	116,002	429,946	37,975	10,421	11,173,049	11,283,924
LFS2003a	1,710,156	8,066,513	1,067,594	71,736	368,467	25,591	5,066	11,182,187	11,297,621
LFS2003b	1,763,434	8,153,945	1,051,023	98,765	292,595	60,218	6,198	11,304,356	11,411,351
LFS2004a	1,723,828	8,255,344	1,011,002	73,286	247,738	29,924	4,856	11,238,187	11,378,217
LFS2004b	1,834,702	8,316,208	1,033,349	82,761	317,428	22,401	1,452	11,531,754	11,630,196
LFS2005a	1,841,573	8,426,520	1,047,161	127,118	449,994	22,458	5,851	11,776,470	11,894,320
LFS2005b	2,088,306	8,761,722	1,002,424	131,904	316,763	56,020	15,217	12,123,370	12,287,798
LFS2006a	2,022,490	8,684,171	1,041,249	82,631	667,467	37,760	6,080	12,345,624	12,437,963
LFS2006b	2,127,488	8,995,048	1,113,504	79,428	463,721	63,927	12,179	12,664,542	12,787,285

(a) Run or do any kind of business, big or small for himself/herself?
(b) Do any work for a wage, salary, commission or any payment in kind?
(c) Do any work as a domestic worker for a wage, salary, or any payment in kind?
(d) Help unpaid in a family business of any kind?
(e) 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?
(f) Do any construction or major repair work on his/her own home, plot, cattle post or business or those of the family?
(g) Catch any fish, prawns, shells, wild animals or other food for sale or family food?
Note: the respondents could answer 'yes' in more than one of the above categories.

Figure A.1: Derivation of employment status (broad definition), OHS1996

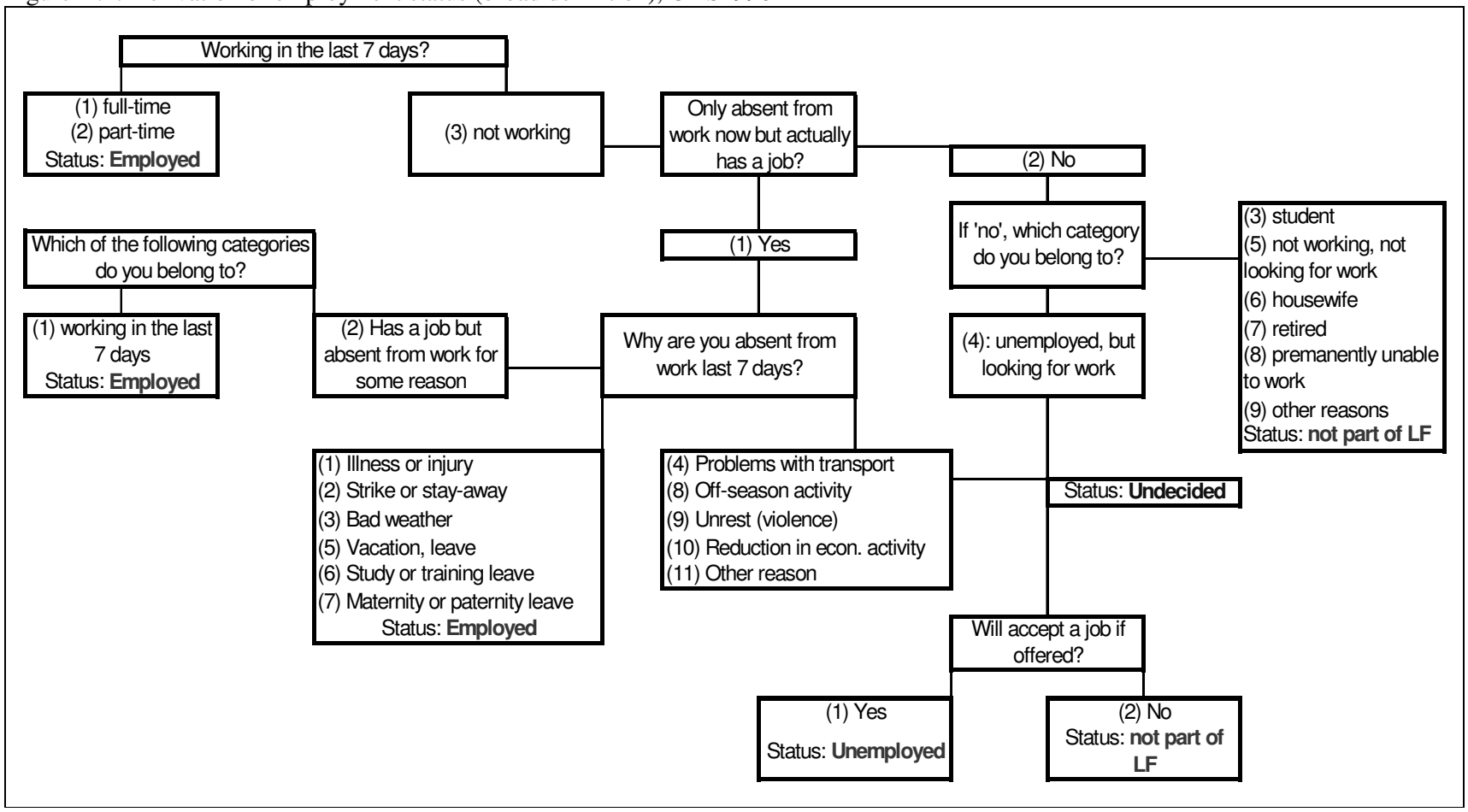


Figure A.2: Derivation of employment status (strict definition), OHS1996

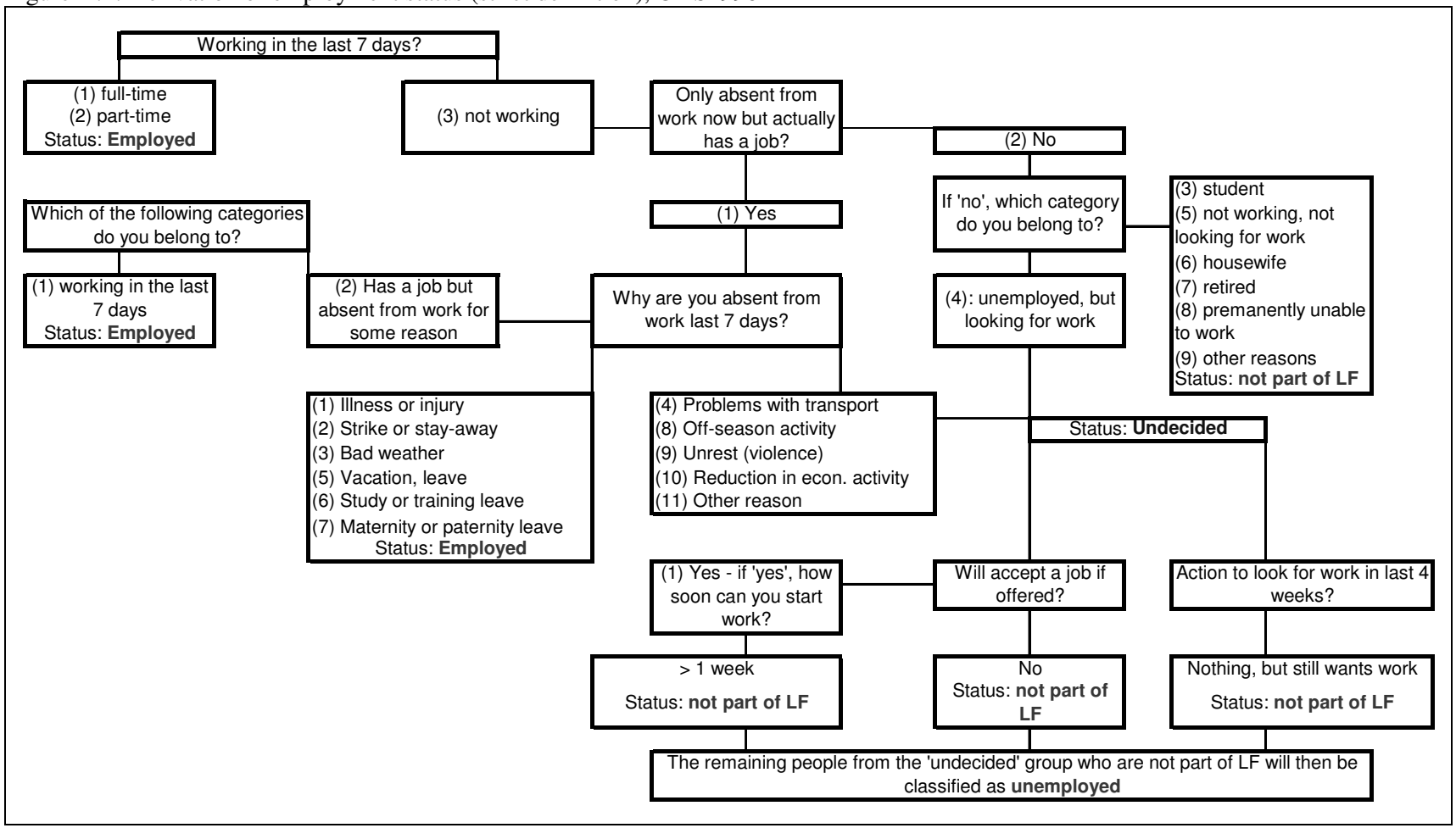


Figure A.3: Derivation of employment status (broad definition), OHS1997 – OHS1998

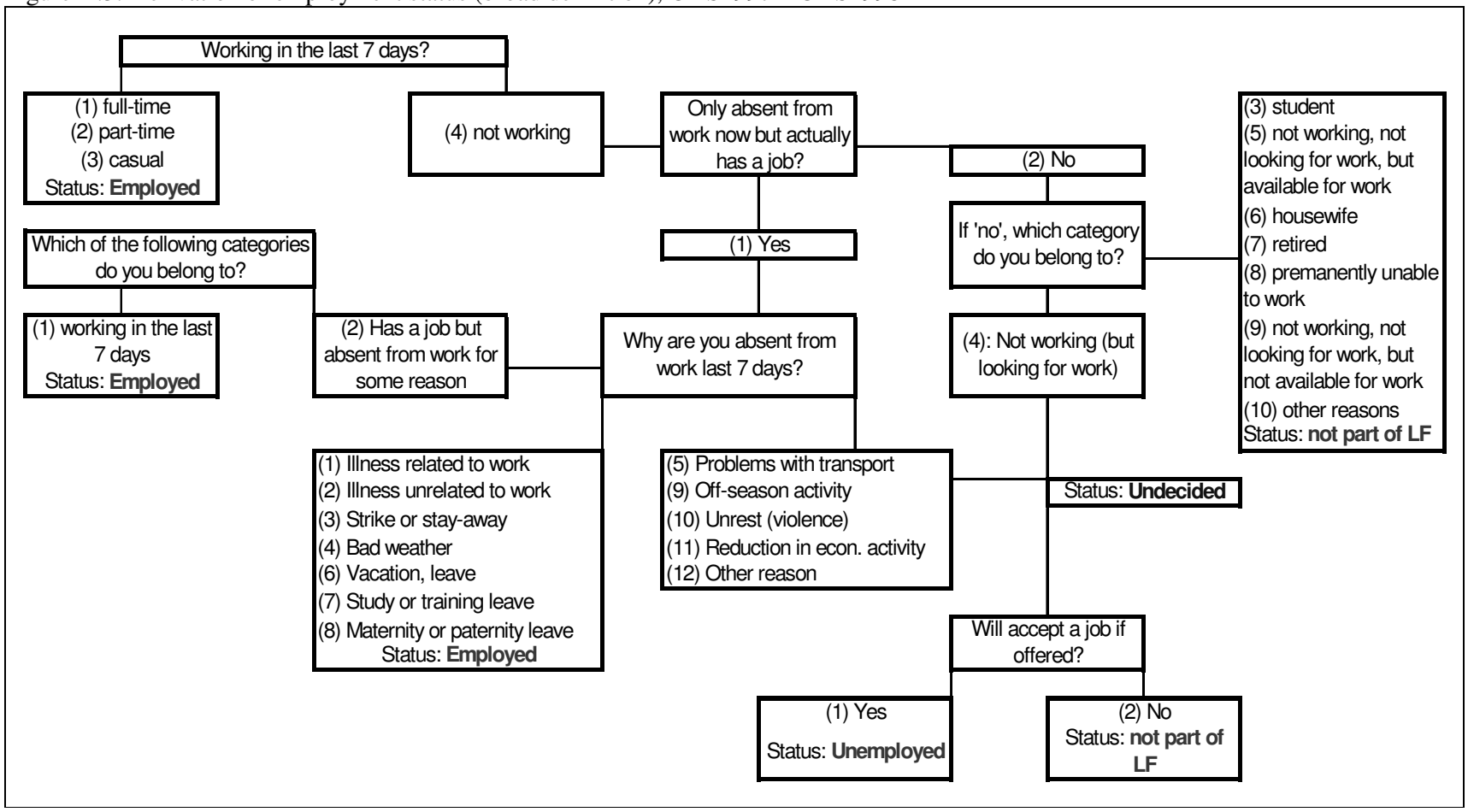


Figure A.4: Derivation of employment status (strict definition), OHS1997 – OHS1998

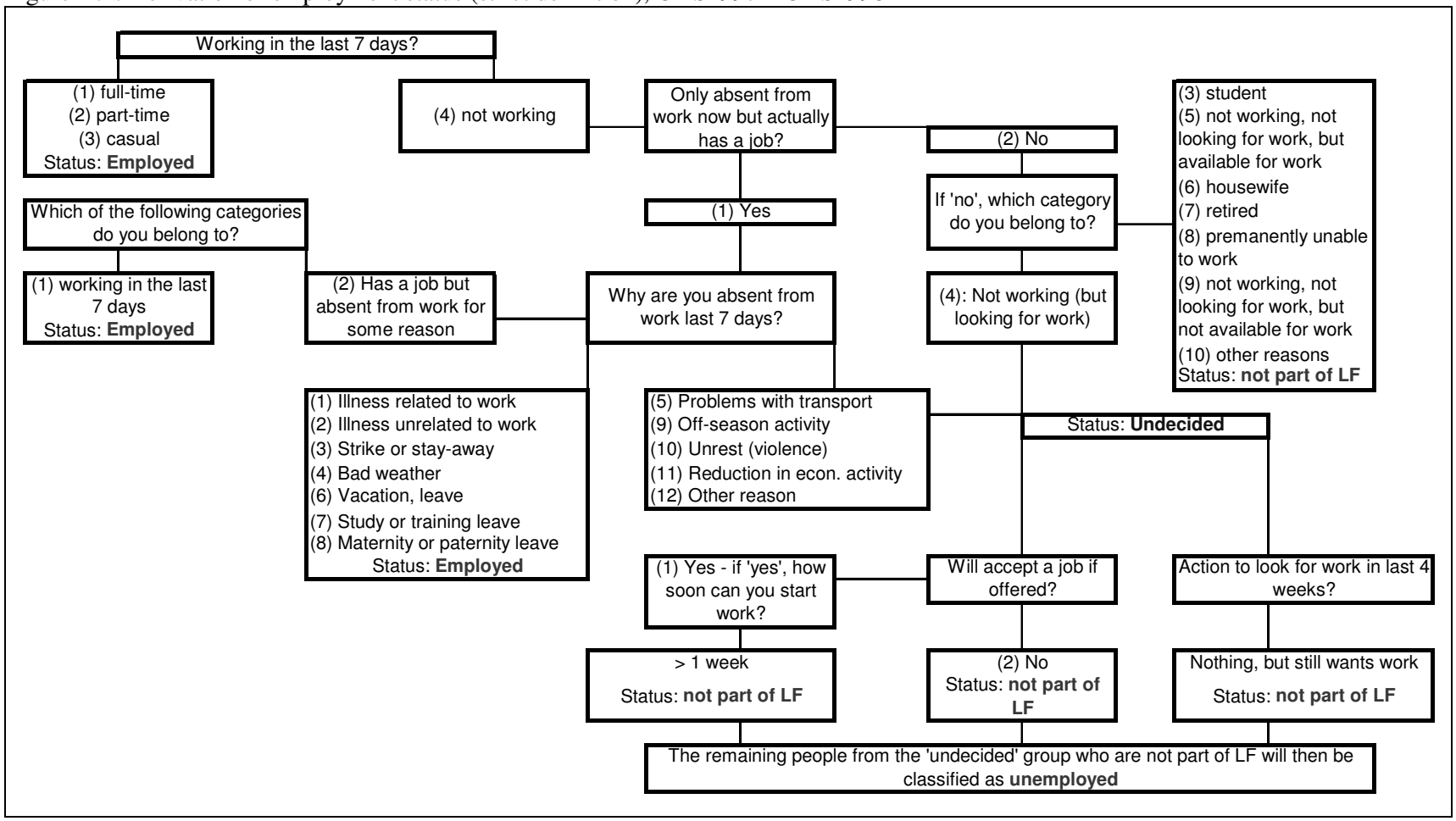


Figure A.5: Derivation of employment status (broad definition), OHS1999

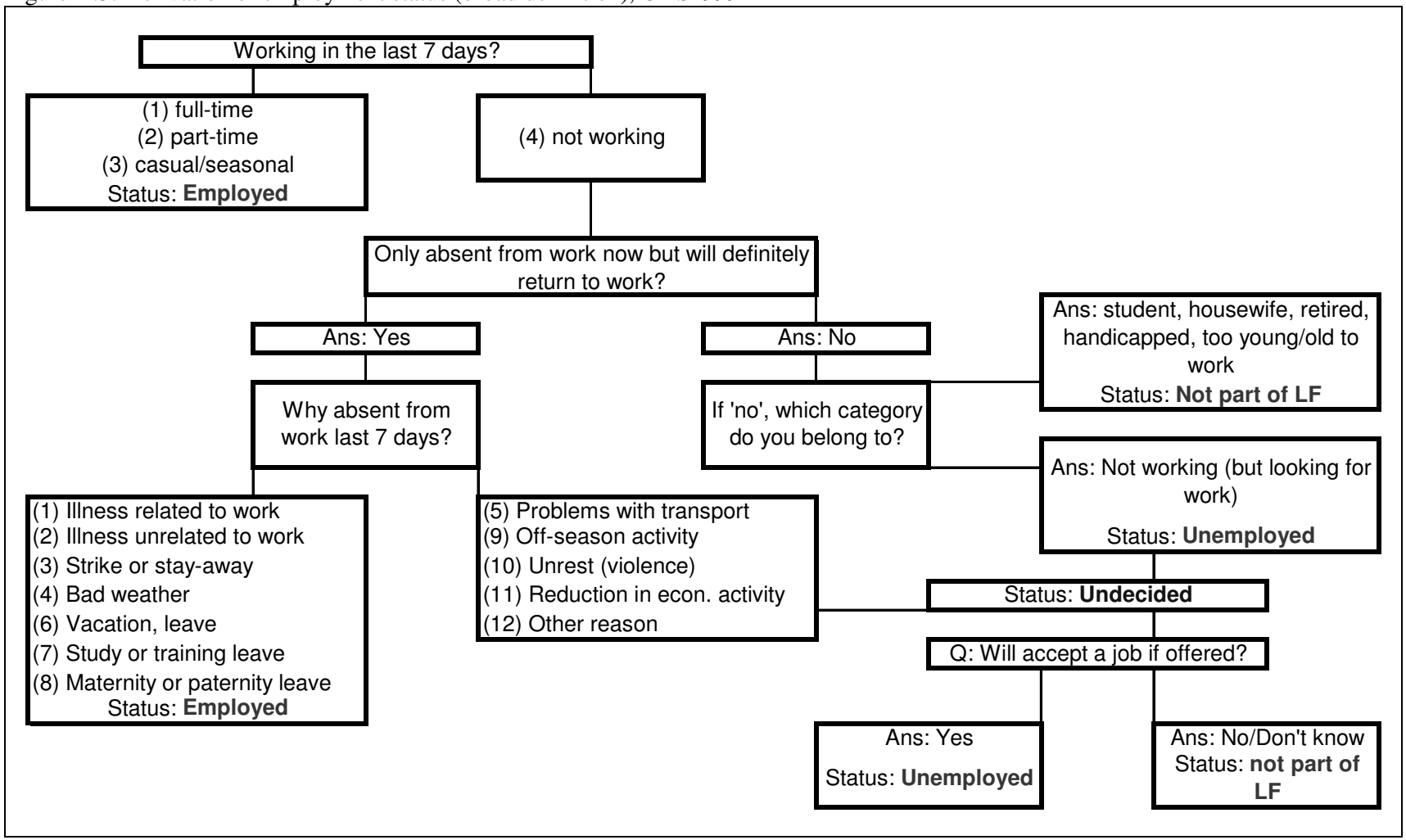


Figure A.6: Derivation of employment status (strict definition), OHS1999

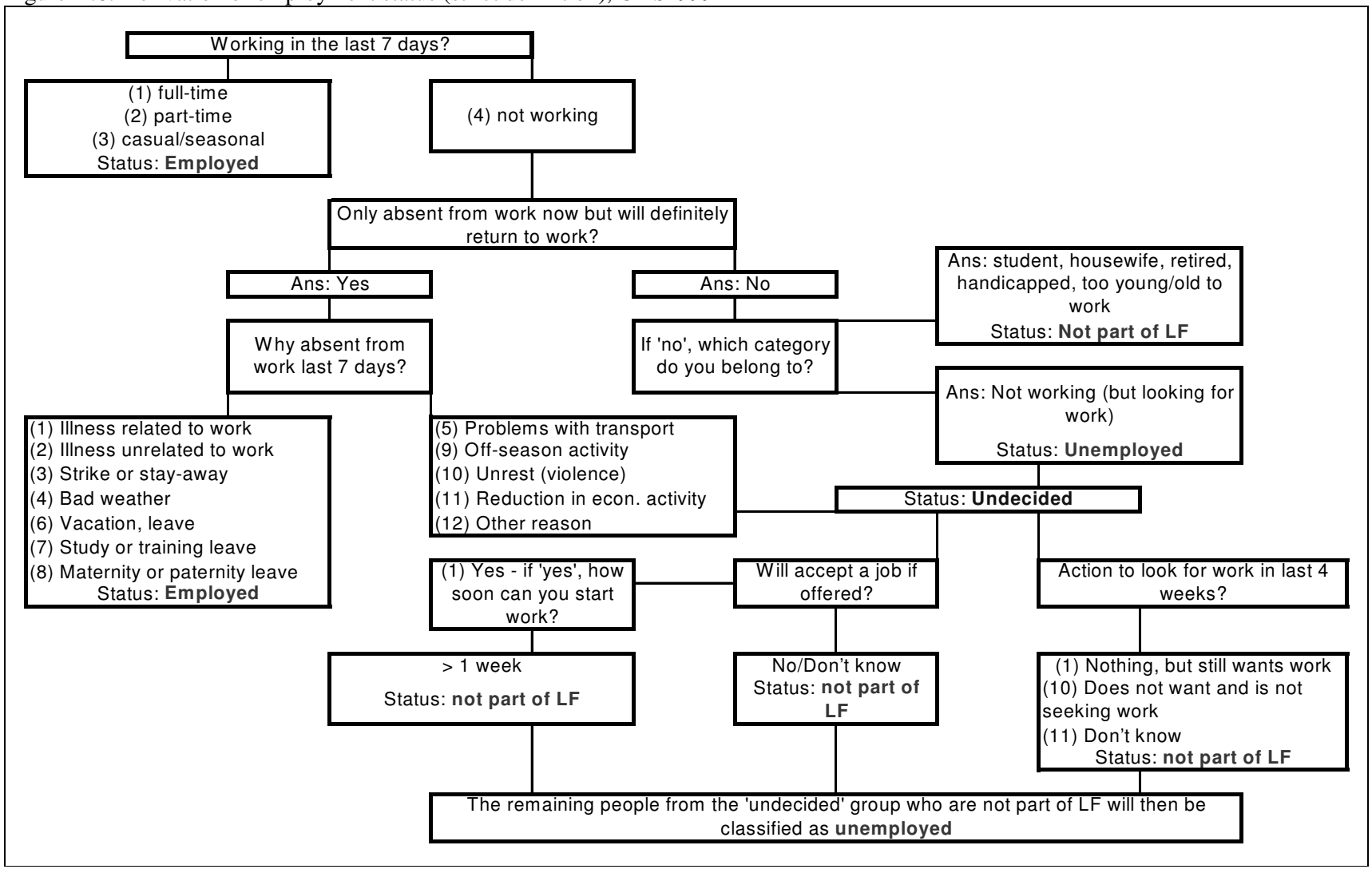


Figure A.7: Derivation of employment status (broad definition), LFS2000a

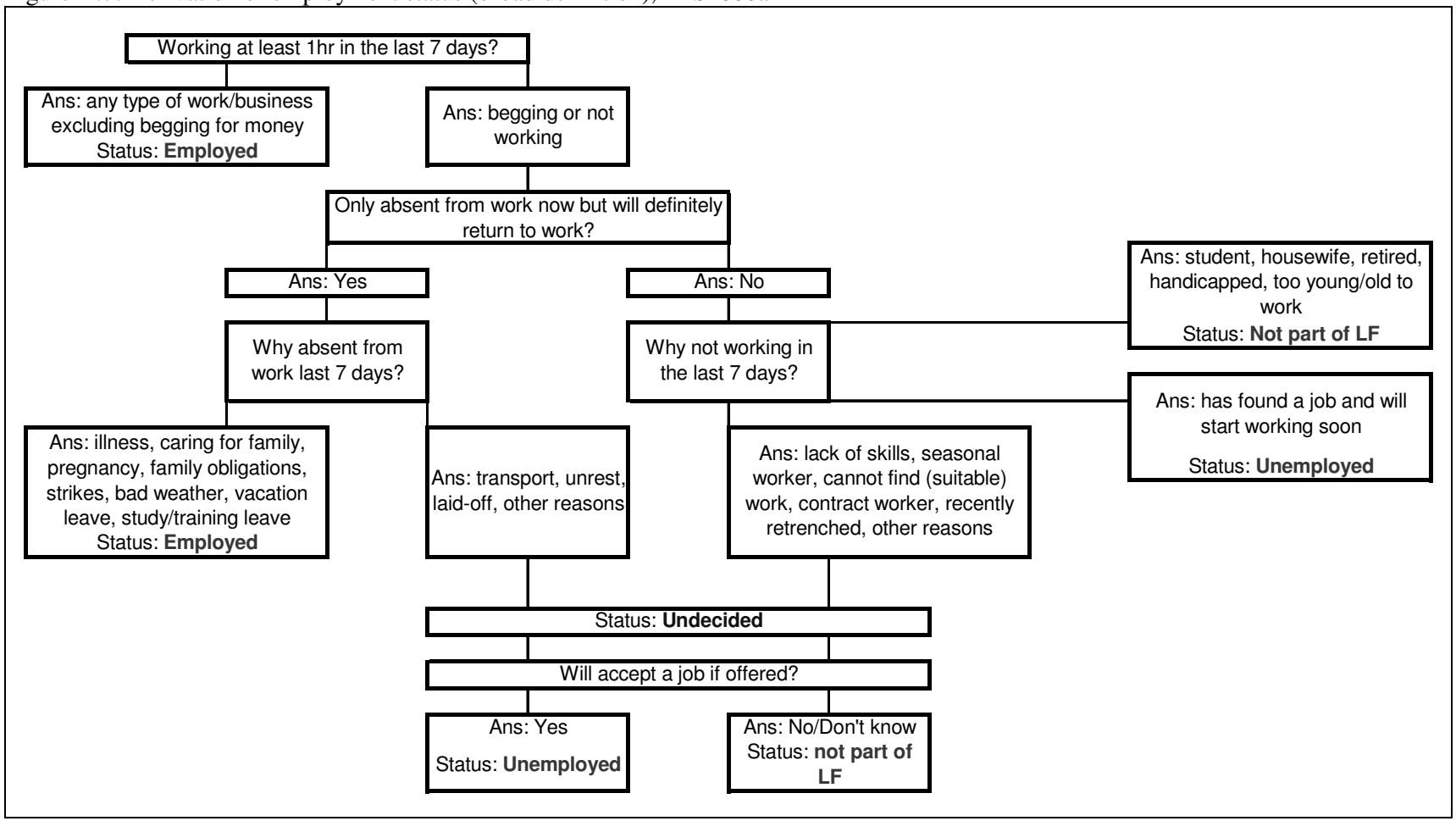


Figure A.8: Derivation of employment status (strict definition), LFS2000a

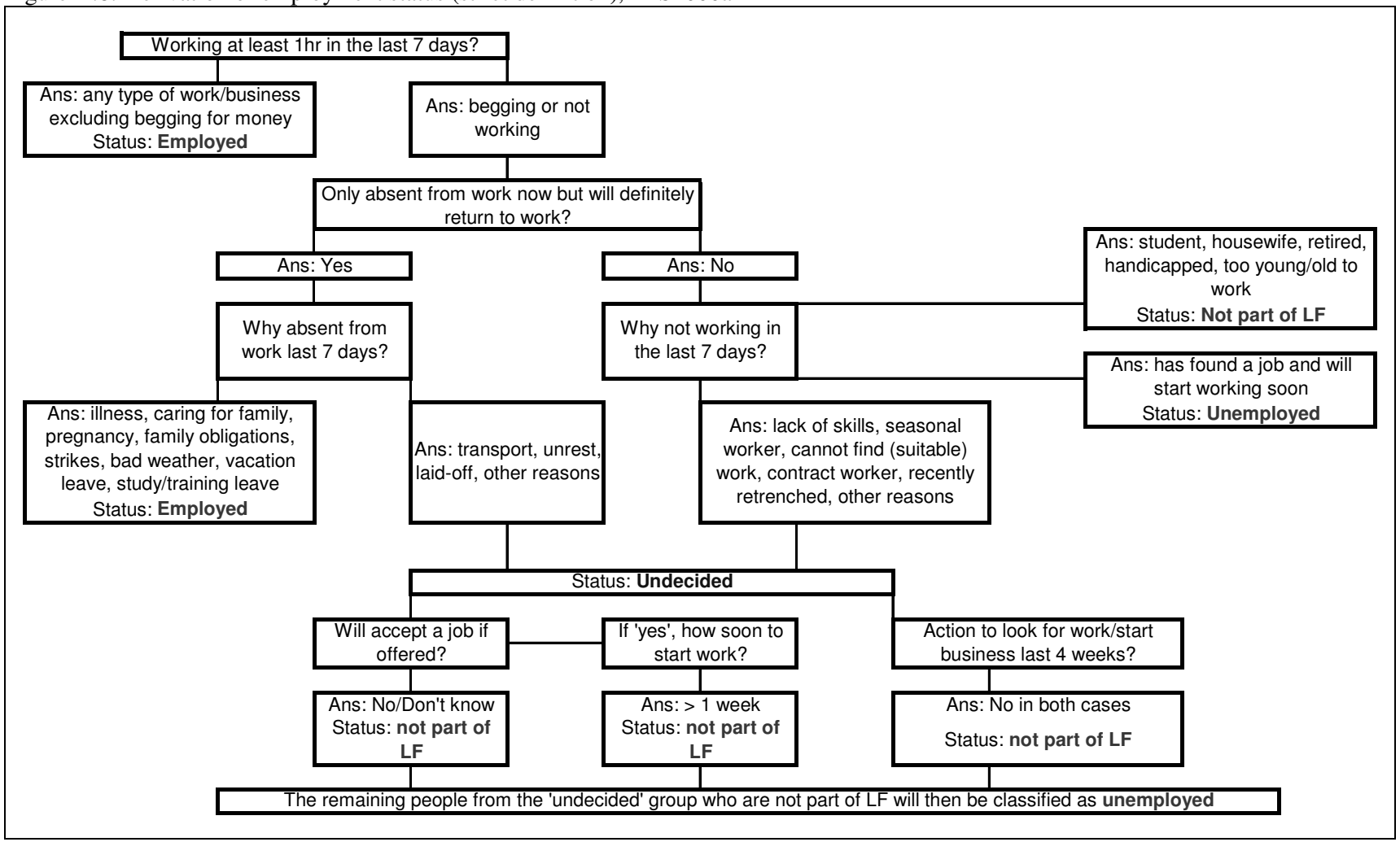


Figure A.9: Derivation of employment status (broad definition), LFS2000b – LFS2006b

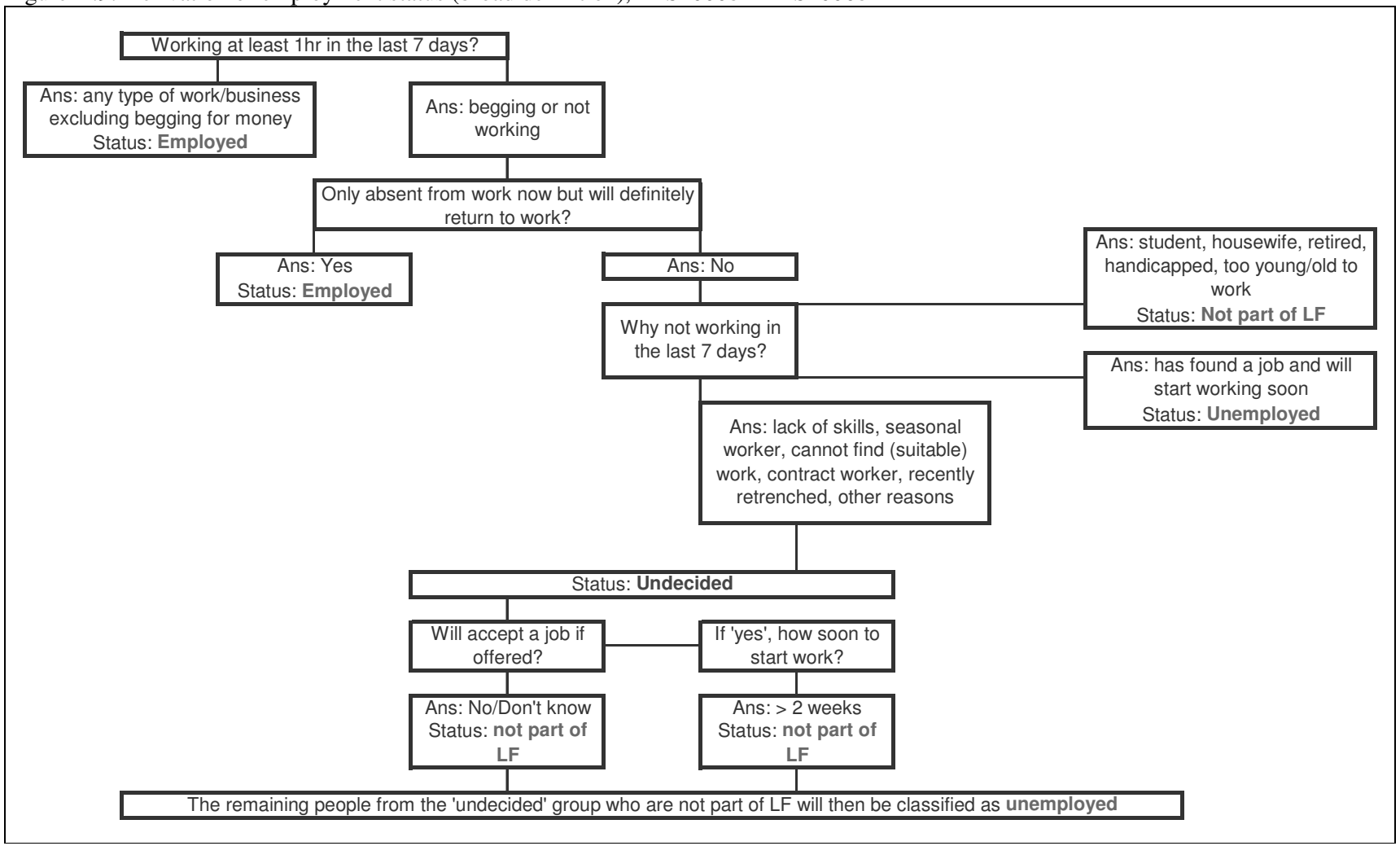


Figure A.10: Derivation of employment status (strict definition), LFS2000b – LFS2006b

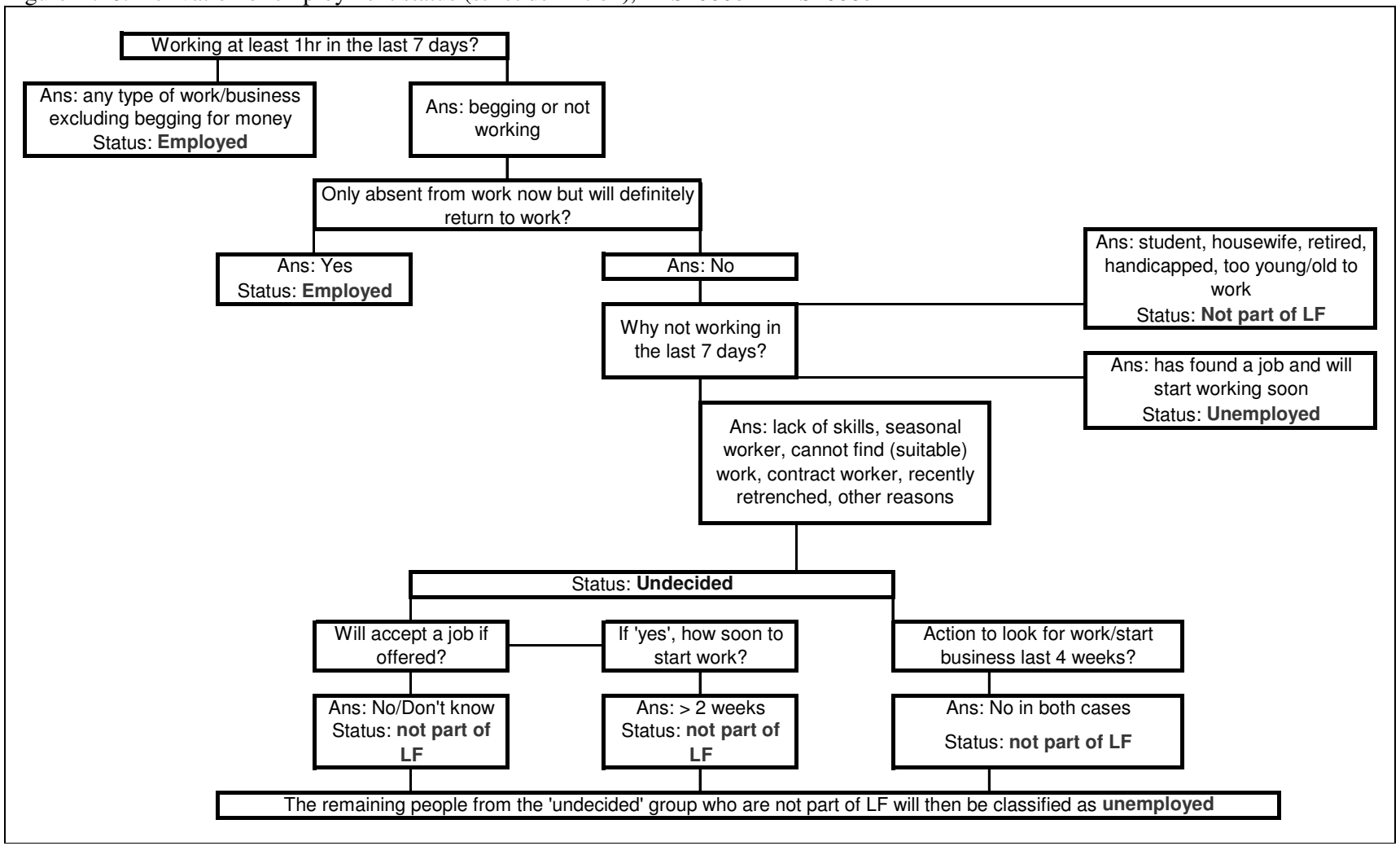


Figure A.11: Derivation of the different categories of formal and informal sector workers