Discussion paper 1
Comparative labour statistics
Labour force survey: first round pilot
February 2000

Statistics South Africa
27 March 2001

DISCUSSION PAPER 1: COMPARATIVE LABOUR STATISTICS

LABOUR FORCE SURVEY: FIRST ROUND PILOT, FEBRUARY 2000

This document presents for discussion a selection of indicative findings and tables from Stats SA's pilot for its first *Labour force survey* (LFS), conducted in February 2000. It examines various labour market issues, including employment in both the formal and informal sectors of the country, and unemployment. The pilot survey provides detailed information regarding approximately 25 000 adults of working age (15 – 65 years) living in 10 000 households across the country. The document also compares available key data related to employment and unemployment in February 2000 with data from the *October household survey* (OHS) of 1999, and the *Survey of total employment and earnings* (STEE) of March 2000.

INTRODUCTION

The LFS is a newly introduced, twice-yearly rotating panel household survey, specifically designed to measure the dynamics of employment and unemployment in the country. It plans to measure a variety of issues related to the labour market, including unemployment rates (official and expanded), according to standard definitions of the International Labour Organisation (ILO). ***For these definitions see Note 1 below.***

A rotating panel sample involves visiting the same dwelling units on a number of occasions (in this instance, five at most), and replacing a proportion of these dwelling units each round (in this instance 20%). New dwelling units are added to the sample to replace those that are taken out. The advantage of this type of design is that it offers the ability to see how the work situation of members of the same dwelling units change over time, while retaining the larger picture of the overall employment situation in the country.

The pilot round of LFS fieldwork took place in February 2000, based on a probability sample of 10 000 dwelling units. Future surveys will take place every six months, using a larger probability sample of 30 000 dwelling units. Rotation will take place before the third round of interviews. One fifth of dwelling units will be replaced at this rotation.

This round is considered to be a pilot because it represented the start of the rotating panel survey in South Africa, and it had a relatively small sample size. In addition, both the sampling methodology and the questionnaire design have changed, when compared with the previous measurements of labour market issues in the October household surveys (OHS). Bridging is therefore required between the old and new methods. This discussion document identifies certain differences between the two surveys (LFS and OHS), and their effect on the measurement of employment and unemployment, as well as some first indications of labour market dynamics.

THE LABOUR MARKET IN FEBRUARY 2000

Labour market trends based on the official definition of unemployment

In Table A, Stats SA gives the overall labour market trends for February 2000, based on the official definition of unemployment (s*ee Note 1 for this definition).* It looks at:

(a) the estimated total number of people in the age category 15 – 65 years (those of working age),
(b) the number of people in this age category, who were not economically active (for example, full-time students, full-time homemakers, retired people and the disabled who are unable to work),
(c) those who were economically active (both the employed and the unemployed according to the official definition of unemployment),
(d) the labour market participation rate (the percentage of all people aged 15-65 years who are economically active), and
(e) the labour absorption rate (the percentage of all aged 15-65 years who are employed) in February 2000.

The table shows that, on the basis of the LFS conducted in February 2000, there were 26,5 million people aged between 15 and 65 years. Among these people:

- 16,2 million were economically active, of whom
  - 11,9 million were employed, and
  - 4,3 million were unemployed.
- In addition, 10,2 million were not economically active, of whom
  - 4,8 million were full-time scholars,
  - 1,0 million were full-time homemakers,
  - 0.9 million were disabled or chronically ill, hence unable to work,

- 0,5 million were either too young or too old to work, and
- 0,3 million were retired.

- The probable official unemployment rate is estimated to be about 26,7%.

TABLE A: LABOUR MARKET TRENDS IN FEBRUARY 2000 BASED ON THE OFFICIAL DEFINITION OF UNEMPLOYMENT		
		(000's)
a	**Total employed**	**11 880**
b	**Total unemployed (official definition)**	**4 333**
c	**Total economically active = a + b**	**16 213**
d	**Total not economically active**	**10 242**
e	**Total aged 15- 65 years = c + d**	**26 454**
f	**Official unemployment rate = b * 100 / c**	**26,7%**
g	**Labour market participation rate = c * 100 / e**	**61,3%**
h	**Labour absorption rate = a * 100 / e**	**44,9%**

LABOUR MARKET DYNAMICS:

Comparing employment trends in February 2000 with those in October 1999

Table B examines the numbers of employed people according to OHS with those found in the new LFS. It suggests ways of taking into account both time differences and seasonal variation.

The first notable difference between the OHS of 1999 and the LFS of February 2000 is the number of people working in agriculture.

- It seems likely that the previous OHSs did not adequately capture small-scale and subsistence farmers, whereas the new LFS, which asks more probing questions concerning this issue, identifies a large group of these informal subsistence and small-scale farmers.
- It is also likely that, in the OHS, the majority of these farmers were identified as not economically active rather than unemployed or employed.
- If one subtracts 1,5 million (in informal farming) from the total of 11,9 million employed people, there are 10,3 million employed people in February 2000, which is similar, after taking sampling error into account, to the 10,4 million in October 1999.

The table also shows a slight decrease in non-agricultural informal sector employment. This phenomenon, after taking sampling error into account, could possibly be explained by seasonal variation. In October 1999, in preparation for the Christmas holiday period, more people may have been working in informal businesses such as street trading than in February 2000.

There is also an increase in domestic work when comparing the OHS of October 1999 with the LFS of February 2000. This could possibly be due to better coverage in the LFS.

Formal sector employment

The increase in formal sector employment between October 1999 and February 2000, as shown in Table C, may be attributable partly to sampling error and partly to possible better coverage of small formal businesses in the LFS, as against the OHS. As noted earlier, the main labour market question is more detailed in the LFS than it was in the previous OHSs.

TABLE B: EMPLOYMENT IN OCTOBER 1999 (OHS) AND FEBRUARY 2000 (LFS)		
Labour market variables	**October 1999**	**February 2000**
	N (000's)	**N (000's) (i)**
Total employed	**10 369**	**11 880**
Among the employed: **Employed in the formal sector (excluding agriculture)**	6 564	6 678
Employed in agriculture	1 099	Commercial: 757 Small-scale/subsistence: 1 508
Employed in the informal sector	1 907	1 821
Employed in domestic service	799	1 001
Employed, sector unspecified	-	115

Comparing the official unemployment rate in February 2000 with that of October 1999
In Table C, the first 'February 2000' column, marked (i), classifies those who said they were homemakers, students, disabled etc., but who said they were available to start work within a week of the interview, and who had taken active steps to seek work in the four weeks prior to the interview, as *not economically active.* **This is according to the official definition of unemployment.**

A main difference between the LFS of February 2000 and the OHS 1999 concerns possible changes in the labour market during this time period. Certain people could have lost their jobs between October 1999 and February 2000, while some of the unemployed could have found work. In addition, new people such as students who had completed their studies could have entered the labour market.

The February 2000 column marked (ii) in Table C reclassifies certain people with regard to their labour market status as it could possibly have been in October 1999, to make the data more comparable with October 1999.

- Those who were unemployed in February 2000 because they were seasonal or casual workers are reclassified as *not economically active.* (If we take these people out of the data set, since we do not know what their status was in October 1999, as indicated in Appendix 1, the unemployment rate becomes 23,2%.)
- Those who were unemployed in February 2000, and who said that they had lost their jobs in the six months prior to the interview, were reclassified as *employed*, since they may have actually been employed in October 1999, but had lost their jobs some time between October 1999 and February 2000 (they had been unemployed for six months or less prior to the February 2000 interview).
- Those unemployed aged 15 to 30 years who were new entrants to the labour market (they had never worked before, and had been looking for work for six month or less) were reclassified as *not economically active* in October 1999, since this was probably their status at that time.

When comparing columns (i) and (ii) of the table, we see that there were approximately 356 000 people who had lost their jobs in the six months prior to the interview of February 2000. The number who were not economically active increases from 10,2 million to 10,7 million. On the other hand, both the number of unemployed (4,3 and 3,5 million) and the unemployment rate (26,7% and 22,5%) decrease.

This is one way in which the data can be compared, but there are others. Further suggested ways of comparing the data across the two data sets (the OHS of October 1999 and the LFS of February 2000) are indicated in APPENDIX I. This appendix also gives earlier figures from previous OHSs, for a more complete comparison.

Thus, by taking the circumstances of individuals into account over time to make the data comparable with OHS 1999, the unemployment rate for February 2000 can be adjusted downwards to 22,5%.

After taking sampling error into account, the unemployment rate of 22,5% in February 2000 is directly comparable with the October 1999 rate of 23,3%.

However, the actual number of unemployed people is higher when comparing the OHS of 1999 and LFS of February 2000, since the labour market participation rate (the economically active as a percentage of all those aged 15 – 65 years) has increased. This means that more people (in actual numbers and proportionally) are saying they are available for work in the LFS than in the OHS. The labour absorption rate (the employed as a percentage of all those aged 15 – 65 years) has also increased. This change is probably due to more in-depth probing in the LFS of February 2000 compared to the OHS of October 1999.

TABLE C: LABOUR MARKET STATISTICS, BASED ON ADJUSTMENTS TO MAKE OHS (OCTOBER 1999) AND LFS (FEBRUARY 2000) DATA SETS COMPARABLE, USING THE OFFICIAL DEFINITION OF UNEMPLOYMENT

	Labour market variables	October 1999	February 2000 (i)	February 2000 (ii)
		N (000s)	N (000s)	N (000s)
a	Total employed	10 369	11 880	12 236
b	Total unemployed (official definition)	3 158	4 333	3 545
c	Total economically active = a + b	13 527	16 213	15 782
d	Total not economically active	12 753	10 242	10 673
e	Total aged 15- 65 years = c + d	26 280	26 454	26 454
	Total reclassified:			
	Not economically active	-	164	595
	Unemployed	-	- 164	- 951
	Employed	-	-	356
f	Official unemployment rate = b * 100 / c	23,3%	26,7%	22,5%
g	Labour market participation rate = c* 100 / e	51,5%	61,3%	59,7%
h	Labour absorption rate = a * 100 / e	39,5%	44,9%	46,3%

AN ALTERNATIVE WAY OF LOOKING AT UNEMPLOYMENT

Table D includes some additional people who could be defined as unemployed, in addition to those included in the official definition. It repeats Table A in column (i) but includes this extra dimension in column (ii).

The columns marked (i) and (ii) in the table are distinguished from each other regarding those actually answering the unemployment questions.

- In the LFS questionnaire, *both* those who said they were *unemployed* and those who said they were *not economically active* (i.e. students, homemakers etc.) were asked to indicate whether or not they:
  (a) want to work and are available to start work within a week of the interview, and
  (b) had taken active steps to look for work or to start some form of self-employment in the four weeks prior to the interview.
- The first 'February 2000' column, marked (i), classifies those who said they were homemakers, students, disabled etc., but who said they were available to start work within a week of the interview, and who had taken active steps to seek work in the four weeks prior to the interview, as *not economically active*.
- The second 'February 2000' column, marked (ii), classifies these same people as *unemployed.*

In previous October household surveys, in common with most other countries, those who said they were not economically active were *not* asked about availability for work, or job seeking behaviour.

In order to be consistent over time and with most other countries, the official unemployment rate is based on classifying those who said they were housewives, students, retired etc. as *not economically active*, as in column (i), even if they had taken some steps to look for work in the four weeks prior to the interview.

Nevertheless, the reader should realise that there were approximately 160 000 people seeking work in the four weeks prior to the interview who classified themselves as homemakers, students, etc. If we classify these people as *unemployed*, the labour market situation changes, as indicated in column (ii).

TABLE D: EMPLOYMENT AND UNEMPLOYMENT IN FEBRUARY 2000 OFFICIAL DEFINITION

	Labour market variables	February 2000	
		N (000's)	
		(i)	(ii)
a	**Total employed**	**11 880**	**11 880**
b	**Total unemployed (official definition)**	**4 333**	**4 496**
c	**Total economically active = a + b**	**16 213**	**16 376**
d	**Total not economically active**	**10 242**	**10 078**
e	**Total aged 15- 65 years = c + d**	**26 454**	**26 454**
f	**Official unemployment rate = b * 100 / c**	**26,7%**	**27,5%**

EMPLOYMENT IN THE FORMAL AND INFORMAL SECTORS BY INDUSTRY, AS INDICATED IN THE LFS

Household surveys indicate that the industry in which people work differs according to whether they are employed in the formal or the informal sector. For example, Table E is based on the LFS of February 2000. It examines the industry in which people work in the both the formal and informal sectors. Domestic work is indicated in a separate column. The table shows that:

- Among people working in the formal sector, 23% were found in community and social services, while 20% were involved in wholesale and retail trade, 17% in manufacturing and 10% in agriculture.
- Among those working in the informal sector, on the other hand, 45% were found in agriculture (the LFS probably measured this type of employment better than the previous OHSs), 29% in trade and 5% in both manufacturing and community services.

TABLE E: EMPLOYMENT IN THE FORMAL AND INFORMAL SECTORS BY INDUSTRY, FEBRUARY 2000*

Industry	Formal		Informal		Domestic		Total*	
	N (000s)	%	N (000s)	%	N (000s)	%	N (000s)	%
Agriculture	757	10,2	1 508	45,3	-		2 285	19,2
Mining	462	6,2	4	0,1	-		467	3,9
Manufacturing	1 277	17,2	178	5,3	-		1 469	12,4
Electricity	86	1,2	2	0,1	-		88	0,7
Construction	388	5,2	196	5,9	-		596	5,0
Trade	1 449	19,5	962	28,9	-		2 434	20,5
Transport	445	6,0	99	3,0	-		547	4,6
Business services	770	10,4	62	1,9	-		837	7,0
Community services	1 724	23,2	158	4,7	-		1 900	16,0
Private households	30	0,4	156	4,7	1001	100,0	1 187	10,0
Other/unspecified industry	46	0,6	3	0,1	-		69	0,6
Total	**7 434**	**100,0**	**3 329**	**100,0**	**1 001**	**100,0**	**11 880**	**100,0**

* The "Total" column includes an extra 87 000 people who did not know whether they worked in the formal or informal sectors

COMPARING FORMAL EMPLOYMENT TRENDS IN THE LFS OF FEBRUARY 2000 WITH THOSE IN THE SURVEY OF TOTAL EMPLOYMENT AND EARNINGS (STEE) OF MARCH 2000

Formal sector employment may be obtained from various other Stats SA data sets, including the *Survey of total employment and earnings* (STEE), which collects information on formal employment in South Africa (excluding certain industries, divisions and activities as described below).

The comparable results to the LFS of February 2000 are from the STEE for March 2000. These were published in July 2000 in *statistical release* P0271. It needs to be borne in mind that STEE obtains data from businesses, while in both the LFS and the OHS, a household rather than a business, is sampled.

Households contain people working in all industries. When complex probability sampling is used, people have the same chance of being selected in their particular stratum in the sample as their overall proportion in a particular industry within that stratum.

The STEE on the other hand collects information from formal sector businesses, *excluding* the following:

- agriculture, hunting, forestry and fishing,
- restaurants and other eating and drinking places, boarding houses, caravan parks, guest farms,
- water and air transport,
- financial institutions,
- real estate and business services,
- educational services outside the public sector,
- medical, dental and other health services outside the public sector,
- welfare organisations outside the public sector,
- religious organisations,
- recreational and cultural services,
- household services and domestic workers in private households, and
- informal businesses.

Table F indicates that STEE shows a decline in formal sector employment in the industries and divisions that it includes. The household surveys (OHS and LFS), on the other hand, show an increase in formal sector employment over time, mainly in the industries and divisions that are not well covered by STEE (see Table F).

The STEE releases generally show a steady decline in formal sector employment, from 5,2 million in September 1996 to 4,7 million in March 2000.

TABLE F: COMPARISONS OF FORMAL SECTOR EMPLOYMENT IN THE OHS OF OCTOBER 1999, THE LFS OF FEBRUARY 2000 AND THE STEE OF SEPTEMBER 1999 AND MARCH 2000

	September/ October 1999	**February/ March 2000**
	(000's)	**(000's)**
Employed according to STEE	4 840	4 754
Employed in formal sector in activities which are not covered in STEE	1 724	1 924

FORMAL SECTOR COMPARISONS BY INDUSTRY IN THE LFS OF FEBRUARY 2000 AND THE STEE OF MARCH 2000

As indicated previously, the STEE excludes a range of businesses not only in certain industries but also in divisions within industries, which may indeed be formal businesses.

Table G indicates that, with regard to numbers of employed people, the only directly comparable industry across STEE and LFS is manufacturing.

STEE, as compared to LFS, finds less employment in business services (574 000 fewer jobs), trade (566 000), community services (235 000), transport (223 000), construction (164 000), electricity (46 000), and mining (41 000).

It would therefore seem as if formal sector employment is growing in those industries or divisions that are not well covered by STEE. This may be indicative of a shift to the tertiary rather than the secondary sector in new job opportunities. It may also be indicative of outsourcing of non-core business in some industries.

TABLE G: THE FORMAL SECTOR AS MEASURED IN THE LFS OF FEBRUARY 2000 AND THE STEE OF MARCH 2000				
	LFS FEBRUARY 2000		STEE MARCH 2000	
	N (000'S)	%	N (000'S)	%
Agriculture	(757)	(excluded)	-	-
Mining	462	7,0	421	8,9
Manufacturing	1 277	19,3	1 278	26,9
Electricity	86	1,3	40	0,8
Construction	388	5,9	224	4,7
Trade	1 449	22,0	883	18,6
Transport	445	6,7	222	4,7
Business services	770	11,7	196	4,1
Community services (excluding domestic)	1 724	26,1	1 489	31,3
Home businesses, other and unspecified	(76)	(excluded)	-	-
Total (excluding agriculture, home business, other and unspecified)	**6 601**	**100,0**	**4 754**	**100,0**

LABOUR MARKET TRENDS, BASED ON THE EXPANDED DEFINITION OF UNEMPLOYMENT

The main difference between the official and the expanded definitions of unemployment is the requirement in the former that, in order to be classified as unemployed, a person must have engaged in job seeking in the four weeks prior to the interview for the survey (*see Note 1 for both definitions).* These criteria have a significant effect on the size of what is considered to be the labour market. Table B below gives overall labour market trends in February 2000 based on the expanded definition of unemployment. In column (i), students, homemakers, the disabled who are unable to work, the retired etc. who are available for work and can start work in the week following the interview are classified as *not economically active*, while in column (ii) they are classified as *unemployed*.

Table H shows that, using the expanded definition, the size of the economically active population, the number of unemployed people and the unemployment rate increase substantially. A large group of people who were available for work did not actively seek work in the four weeks prior to the February interview.

TABLE H : LABOUR MARKET STATISTICS BASED ON THE EXPANDED DEFINITION OF UNEMPLOYMENT, FEBRUARY 2000			
Labour market variables		February 2000	
		N (000's)	
		(i)	(ii)
A	**Total employed**	**11 880**	**11 880**
B	**Total unemployed (expanded definition)**	**6 553**	**7 075**
C	**Total economically active = A + B**	**18 432**	**18 955**
D	**Total not economically active**	**8 022**	**7 499**
E	**Total aged 15- 65 years = C + D**	**26 454**	**26 454**
F	**Expanded unemployment rate = B* 100 / C**	**35,5%**	**37,3%**

Breakdown of unemployment rates by urban and non-urban areas, gender, and race
Table I below compares unemployment rates by urban or non-urban place of residence, race and gender, in terms of the *official* definition of unemployment, while Table J compares them using the *expanded* definition.

Table I shows that:

- Unemployment rates were higher in urban, as against non-urban areas, probably because of good coverage of subsistence and small-scale agriculture.
- The unemployment rate is highest among Africans and lowest among whites.
- It is higher among women than men.

TABLE I: OFFICIAL UNEMPLOYMENT RATES AMONGST MALES AND FEMALES LIVING IN URBAN AND NON-URBAN AREAS BY POPULATION GROUP, FEBRUARY 2000

Gender, population group	Urban male	Urban female	Non-urban male	Non-urban female	Total male	Total female	Total
(i)	%* (ii)	%* (iii)	%* (iv)	%* (v)	%* (vi)	%* (vii)	%* (viii)
(a) All population groups:	26,1	31,2	22,8	24,4	24,9	28,7	26,7
(b) African:	34,2	40,2	24,6	25,1	30,0	33,2	31,6
(c) Coloured:	22,6	22,9	5,9	13,7	19,5	21,4	20,4
(d) Indian:	16,6	24,6	--**	--**	16,7	24,8	19,9
(e) White:	6,2	7,3	--**	17,4	5,9	7,9	6,8

* Each percentage is a percentage of all people in that particular category.
** Number of responses were too few for this analysis

Table J shows the corresponding breakdowns for the expanded unemployment rate. It shows a similar pattern to Table I, but, as can be expected, the unemployment rates are higher.

TABLE J: EXPANDED UNEMPLOYMENT RATES AMONGST MALES AND FEMALES LIVING IN URBAN AND NON-URBAN AREAS BY POPULATION GROUP, FEBRUARY 2000

Gender, population group	Urban male	Urban female	Non-urban male	Non-urban female	Total male	Total female	Total
(i)	%* (ii)	%* (iii)	%* (iv)	%* (v)	%* (vi)	%* (vii)	%* (viii)
(a) All population groups:	31,0	39,3	33,3	39,3	31,9	39,3	35,5
(b) African:	39,8	48,2	35,5	40,4	37,8	44,4	41,2
(c) Coloured:	27,6	31,3	9,8	22,3	24,3	29,8	27,1
(d) Indian:	19,0	37,8	--**	--**	19,1	38,0	27,3
(e) White:	8,2	12,2	--**	18,2	8,2	12,6	10,1

* Each percentage is a percentage of all people in that particular category.
** Number of responses were too few for this analysis

Mr Pali Lehohla
Statistician General: Statistics South Africa

NOTES

1. Official and expanded unemployment rates

Statistics South Africa (Stats SA) uses the following definition of unemployment as its *official* definition. The *unemployed* are those people within the *economically active population* who: (a) did not work during the seven days prior to the interview, (b) want to work and are available to start work within a week of the interview, and (c) have taken active steps to look for work or to start some form of self-employment in the four weeks prior to the interview. The *expanded unemployment rate* excludes criterion (c).

Among those who are included in the expanded but not the official definition of unemployment will be discouraged job seekers (those who said they were unemployed but had not taken active steps to find work in the four weeks prior to the interview).

Stats SA reports on the situation of the unemployed using both the official and the expanded definition. In the present economic climate, there is a proportion of discouraged work seekers who face constraints, for example high travel costs and lack of transport, when seeking work.

2. Sample design for the LFS of February 2000

The OHSs were independent cross-sectional surveys, and different samples were designed for each of them. For the LFSs, on the other hand, a rotating panel sample design is being used, to allow for measurement of change in people's employment situation over time. The same dwellings will be visited on, at most, five different occasions. After this, new dwelling units will be included for interviewing from the same PSU in the master sample. This means a rotation of 20% of dwelling units each time.

The LFS sample of February 2000 was drawn from a master sample, in which the same primary sampling units will be visited in future LFSs. The database of enumerator areas (EAs), as established during the demarcation phase of Census '96, constituted the sampling frame for selecting EAs for the LFS. As part of the master sample, small EAs consisting of fewer than 100 dwelling units were combined with adjacent EAs to form primary sampling units (PSUs) of at least 100 dwelling units, to allow for repeated sampling of dwelling units within each PSU. The sampling procedure for the master sample involved explicit stratification by province and within each province, by urban and non-urban areas. Independent samples of PSUs were drawn for each stratum within each province. The smaller provinces were given a disproportionately larger number of PSUs than the bigger provinces.

Altogether, 3 000 PSUs were drawn for the master sample, by means of probability proportional to size principles in each stratum. The measure of size was the number of dwelling units in each PSU. A subset of 1 000 PSUs was drawn for the pilot LFS of February 2000. In October 2000, all 3 000 EAs will be used to have a sample of 30 000 dwelling units. Interviewing for the February 2000 LFS took place in these 1 000 PSUs, and simple random sampling was applied to select 10 dwelling units to visit in each PSU as ultimate sampling units.

3. Weighting the LFS of February 2000

The February 2000 LFS was weighted to estimates of the population size based on the population census of October 1996, as adjusted by a post-enumeration survey (PES), using post-stratification by province, gender and five-year interval age groups.

4. Symbols used in the tables that follow

When a zero (0) is shown in a table, there were fewer than 500 respondents, after weighting, in this category.
When a dash (-) is shown there were no respondents in the category.
When a single asterisk (*) is shown in the table, the sample size was too small to give reliable estimates.

5. Comparability of results with other Stats SA data sources

The *Survey of total employment and earnings* (STEE) collects information on formal employment in South Africa. The comparable results of the STEE, i.e. for March 2000, were published in July 2000 in *statistical release* P0271.

6. Urbanisation

The urban population constituted 54,1% of the total population according to *Census '96*. In the weighting matrix for the LFS of February 2000, the proportionate distribution of the population by urban and non-urban areas was based on the population census of 1996. It follows that urbanisation cannot be detected from the LFS, but will be measured by comparing *Census '96* with *Census 2001*.

7. Confidence intervals

Stats SA have calculated 95% confidence limits for some key variables. These are available on request to users who require this information.

DEFINITIONS OF TERMS

A *household* consists of a single person or a group of people who live together for at least four nights a week, who eat together and who share resources.

Population group describes the racial classification of a particular group of South African citizens. The previous government used legislation to impose this type of classification, to divide the South African population into distinct groupings on which to base apartheid policies. For quite a different reason it remains important for Stats SA to continue to use this classification wherever possible. It clearly indicates the effects of discrimination of the past, and permits monitoring of policies to alleviate discrimination. Note that, in the past, population group was based on a legal definition, but it is now based on self-perceptions and self-classification. An *African/black* person is someone who classifies him/herself as such. The same applies to a *coloured, Indian/Asian or white* person.

A hostel is a communal living quarter for workers, provided by a public organisation such as a local authority, or a private organisation such as a mining company. These were residential dormitories established for migrant workers during the apartheid era, and they continue to house people working in certain industries, such as the mining industry.

Institutions are communal temporary, semi-permanent or permanent living arrangements for people in special circumstances, for example prisons, police cells, school boarding facilities, homes for the aged or the disabled, hotels and hospitals.

The *working age population* includes all those aged between 15 and 65 years.

The *economically active population* consists of both those who are employed and those who are unemployed.

The *employed* are those who performed work for pay, profit or family gain in the seven days prior to the household survey interview, or who were absent from work during these seven days, but did have some form of paid work during this time.

The *official unemployment rate:* see Note 1.

The *expanded unemployment rate*: see Note 1.

The people who are *out of the labour market* or who are *not economically active* are those who are not available for work. This category includes full-time scholars and students, full-time homemakers, those who are retired, and those who are unable or unwilling to work.

The formal sector includes all businesses which are registered for tax purposes, and which have a VAT number.

The informal sector consists of those businesses that are unregistered and do not have a VAT number. They are generally small in nature, and are seldom run from business premises. Instead, they are run from homes, street pavements or other informal arrangements.

Primary industries include agriculture, forestry and fishing, and mining and quarrying.

Secondary industries include manufacturing, electricity and other utilities, and construction.

Tertiary industries include trade, transport, financial and business services, and social, personal and community services.

Employment status refers to whether or not the person is self-employed, or works as an employee, or both, or else works as a domestic worker in a household.

Location refers to whether the person lives in an urban or non-urban area.

- An *urban* area is one that has formally been legally proclaimed as being urban. These include towns, cities and metropolitan areas.
- A *semi-urban* area is not part of a legally proclaimed urban area, but adjoins it. Informal settlements are examples of these types of areas. In this publication *semi-urban* areas have been *included* with non-urban areas.
- All other areas are classified as *non-urban*, including commercial farms, small settlements, rural villages and other areas which are further away from towns and cities.

Workers include the self-employed, employers and employees.

Labour market dynamics refer to movement into and out of the labour market, and into and out of actual employment, over a specified time period.

APPENDIX I

Appendix 1 compares labour market statistics from consecutive OHSs from 1996 to 1999 and the first pilot LFS. It classifies, except where indicated, those who were available for work, who could start work in the week after the interview, and who had looked for work in the four weeks prior to the interview as being unemployed.

- Column (i), on the left-hand side, indicates the variables that are being compared.
- Column (ii) gives the data for the OHS of 1996.
- Column (iii) gives the data for the OHS of 1997.
- Column (iv) gives the data for the OHS of 1998.
- Column (v) gives the data for the OHS of 1999.
- Column (vi) gives the data for February 2000, with homemakers, students etc. who had looked for work in the four weeks prior to the interview classified as *unemployed*.
- Column (vii), ***based on the official definition of unemployment***, gives the data for February 2000 (homemakers, students etc. who had looked for work in the four weeks prior to the interview classified as *not economically active).*
- Column (viii) gives the data for February 2000, but it changes the classification of certain individuals to take labour market dynamics into account. Homemakers, students etc. who had looked for work in the four weeks prior to the interview are classified as *not economically active.* In addition, *seasonal workers* and those who had entered the labour market in the *month* prior to the interview are also classified as *not economically active*. Those who had lost their jobs in the *month* prior to the interview are classified as *employed.* This is the first suggestion to make the data comparable with October 1999.
- Column (ix) gives the data for February 2000, with homemakers, students etc. who had looked for work in the four weeks prior to the interview classified as *not economically active*. In addition, those who had entered the labour market in the *six months* prior to the interview are also classified as *not economically active*. Those who had lost their jobs in the *six months* prior to the interview are classified as *employed*. This is the second suggestion to make the data comparable with October 1999.
- Column (x) gives the data for February 2000, with homemakers, students etc. who had looked for work in the four weeks prior to the interview classified as *not economically active*. In addition, those who had entered the labour market in the six months prior to the interview are also classified as *not economically active*. *Seasonal workers in February 2000 are also re-classified as not economically active in October 1999.*Those who had lost their jobs in the six months prior to the interview are classified as *employed*. This is the third suggestion to make the data comparable with October 1999.

Unfortunately, the questionnaire did not break down the time of unemployment between one and six down into smaller time-based categories for measuring change in status more precisely, when looking at labour market dynamics.

When comparing columns (vii) and (viii) of Appendix 1, for example, we see that there were approximately 57 000 people who lost their jobs in the month prior to the interview of February 2000. Thus they were employed in October 1999. The number who were not economically active increased from 10,2 million to 10,4 million, since a group of those who were students in 1999 could have entered the labour market in February 2000. The number of unemployed (4,3 and 4,1 million) and the unemployment rate (26,7% and 25,6%) also change.

The second column from the right marked February 2000 (ix) in Appendix A shows an alternative way of comparing data between October 1999 and February 2000. It reclassifies new job seekers entering the labour market some time in the six months prior to the interview (rather than one month, as in (viii)), to not economically active people in October 1999. It also reclassifies those who have lost jobs some time in the past six months to being employed in October 1999. The casual/ seasonal worker reclassifications used in (viii) are *excluded.* The unemployment rate reduces to 23,2%.

The right-hand February 2000 column marked (x) is calculated in the same way as column (viii), with regard to their labour market status. But in this instance, the criterion of being unemployed for six months or less prior to the interview is applied, as against the criterion of one month or less in (viii).

APPENDIX 1: LABOUR MARKET STATISTICS, INCLUDING THE MINING SECTOR, BASED ON THE OFFICIAL DEFINITION OF UNEMPLOYMENT, OCTOBER 1996, 1997, 1998 AND 1999										
		1996	**1997**	**1998**	**1999**	**2000**	**2000**	**2000**	**2000**	**2000**
Labour market variables (i)		**N (000's) (ii)****	**N (000's) (iii)****	**N (000's) (iv)****	**N (000's) (v)****	**N (000's) (vi)****	**N (000's) (vii)****	**N (000's) (viii)****	**N (000's) (ix)****	**N (000's) (x)****
a	**Total employed**	**9 287**	**9 247**	**9 390**	**10 369**	**11 880**	**11880**	**11 937**	**12 267**	**12 236**
	Among the employed: Employed in the formal sector STEE survey (excluding agriculture and certain activities not covered in STEE)	5 242	5 139	4 945	4 840	4 754	*4 754*	*4 754*	4 754	4 754
	Employed in agriculture	759	717	935	1 099	F : 757 I: 1 508	*F : 757* *I: 1 508*	*F : 757* *I: 1 508*	F : 757 I: 1 508	F : 757 I: 1 508
	Employed in the formal sector in activities not covered in STEE	1 550	1 587	1 445	1 724	1 924	*1 924*	*1 924*	1 924	1 924
	Employed in the informal sector	* 996	1 136	1 316	1 907	1 821	*1 821*	*1 821*	1 821	1 821
	Employed in domestic service	740	668	749	799	1 001	*1 001*	*1 001*	1 001	1 001
	Employed sector unspecified/unknown	-	-	-	-	115	*115*	*115*	115	115
	Previously employed but lost job	-	-	-	-	-	-	*57*	387	356
b	**Total unemployed (official definition)**	**2 224**	**2 451**	**3 163**	**3 158**	**4 496**	***4 333***	***4 101***	**3 697**	**3 545**
c	**Total economically active = a + b**	**11 511**	**11 698**	**12 553**	**13 527**	**16 376**	***16 213***	***16 037***	**15 964**	**15 782**
d	**Total not economically active**	**13 717**	**13 961**	**13 157**	**12 753**	**10 078**	***10 242***	***10 417***	**10 490**	**10 637**
e	**Total aged 15- 65 years = c + d**	**25 228**	**25 659**	**25 710**	**26 280**	**26 454**	***26 454***	***26 454***	**26 454**	**26 454**
f	**Official unemployment rate = b * 100 / c**	**20,3%**	**22,0%**	**25,2%**	**23,3%**	**27,5%**	***26,7%***	***25,6%***	**23,2%**	**22,5%**

* The questionnaire in 1996, while differentiating between the self-employed and employers in the formal and informal sectors, did not distinguish between *employees* in the formal and those in the informal sectors. The size of the sector in 1996 has therefore been estimated on the basis of the proportions of informal employers and employees in 1997, 1998 and 1999.
** Weights were not scaled for age distribution and gender.
*** Figures for 1996 and 1997 have been revised on the basis of re-weighting in relation to the mining sector.

TECHNICAL NOTE ON ESTIMATION AND USE OF STANDARD ERRORS

The published results of the labour force survey are based on representative probability samples drawn from the South African population, as discussed in the section on sample design. Consequently, all estimates are subject to sampling variability. This means that the (sample) estimates may differ from the figures (i.e. population figures) that would have been produced if the entire South African population had been included in the survey. The measure usually used to indicate the likely difference between a sample estimate and the corresponding population figure is the *standard error (SE)*, which measures the extent to which an estimate might have varied by chance because only a sample of the population was included. There are about two chances in three that the sample estimate will differ by less than one standard error from the population figure and about 19 chances in 20 that the difference will be less than two standard errors. Another measure of the likely difference is the *relative standard error* (or the *coefficient of relative variation, or CV*) which is defined as the standard error of the estimate divided by the size of the estimate, and is usually expressed as a percentage.

There are two major factors which influence the value of a standard error. The first factor is the sample size. Generally speaking, the larger the sample size, the more precise (accurate) the estimate and the smaller the standard error. Consequently, in a national household survey such as the LFS, one expects more precise estimates at the national level than at the provincial level due to the larger sample size involved. The second factor is the variability between households of the parameter of the population being estimated, for example the number of unemployed persons in the household.

For every survey, Statistics South Africa now calculates the standard errors and relative standard errors for a variety of the estimates shown in its publications. Estimates are calculated, not only of various population parameters but also for the many subclasses of the country, which include segregated classes (e.g. explicit strata, such as provinces or urban/rural division or combinations of these) as well as cross-classes (e.g. gender, age groups, gender by age groups). These different subclasses represent a large variety of sample sizes. Smoothing of the calculated standard errors is obtained by fitting regression models to the relative standard errors, which are then represented in graphical form. Given the size of the estimate and the population parameter under consideration, an approximate value of the relative standard error of the estimate can be obtained (read off) from the relevant graph. Multiplication of this approximate value of the relative standard error with the estimate itself then gives an approximate value of the SE of the estimate, viz.

$$(1) \qquad SE(estimate) = CV(estimate) \times estimate$$

The formula in this form, however, is not applicable to ratio estimates, such as the unemployment rate, and has to be changed to:

$$(2) \quad SE(ratio\ \text{estimate}) = CV(ratio\ \text{estimate}) \times (\text{numerator of ratio estimate})$$

For example:

$$(3) \quad \text{SE(unemployment rate)} = \text{CV(unemployment rate)} \times (\text{estimated number of unemployed persons})$$

Note that there are different graphs to be used for the different population parameters for obtaining the CV estimate.

95% confidence intervals of a population parameter can be obtained as follows:

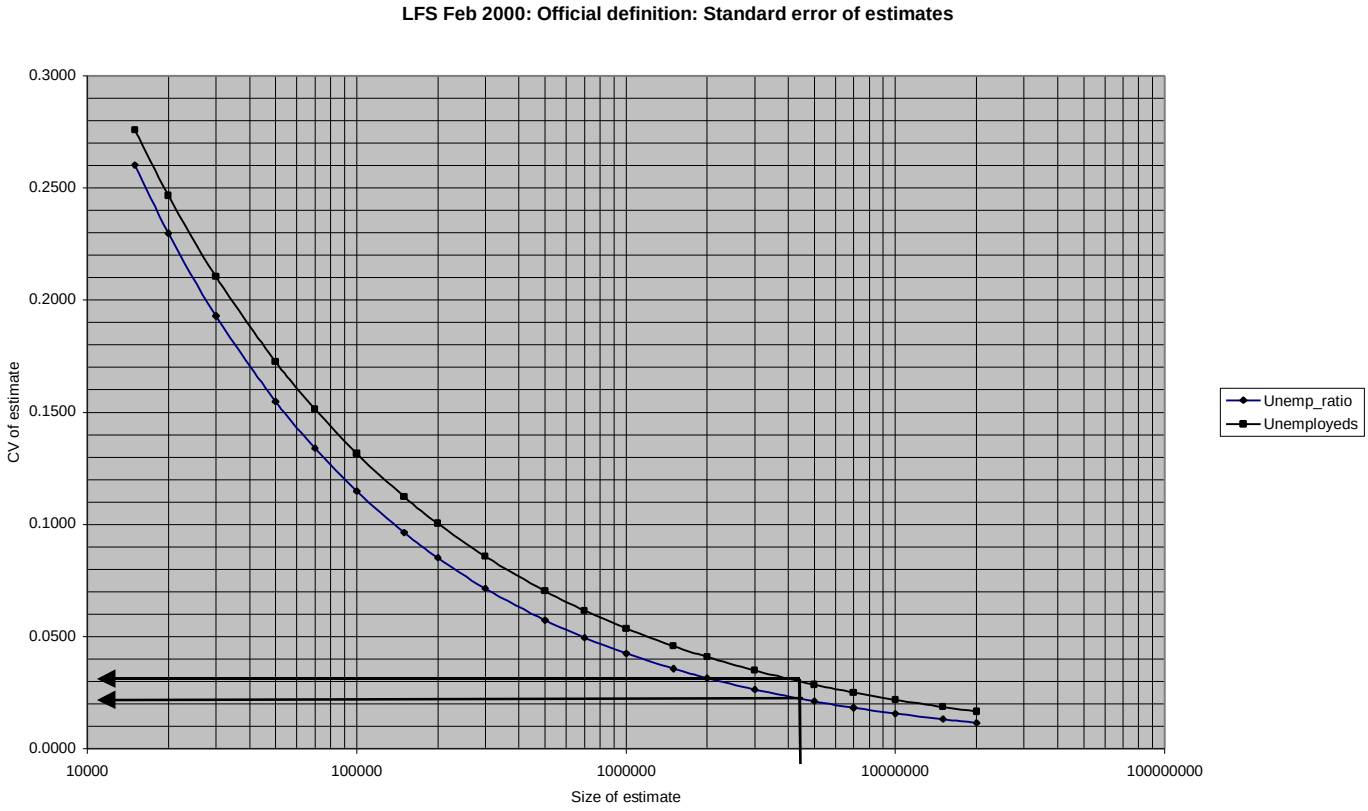
*Lower 95% confidence limit of a population parameter = estimate – 1.96*SE(estimate)*
and
*Upper 95% confidence limit of a population parameter = estimate + 1.96*SE(estimate).*

The *confidence coefficient*, i.e. 95%, of a 95% confidence interval of a population parameter could be interpreted as the success rate of the calculated confidence interval, viz.

*{estimate – 1.96*SE(estimate); estimate + SE(estimate)}*

to include or to contain the value of the population parameter.

Example:



Calculating the standard error of the unemployed according to the official definition. Note that the estimated number of unemployed is 4 333 000. Mark this on the graph and read off the corresponding coefficient of variation.. In our case it is 0.03 and 0.021 for the unemployed and unemployment rate respectively.

Applying formula (1) the standard error for the unemployed will be 0.03x4333=130.

Applying formula (2) the standard error for the unemployment rate will be 0.021x26.7=0.56

This implies that the 95% confidence intervals of the unemployed and the unemployment rate are 4333 ± (1.96x130) and 26.7 ± (1.96x0.56) respectively.