Earnings inequality and unemployment in South Africa

Fiona Tregenna*

Department of Economics and Econometrics, University of Johannesburg, Johannesburg, South Africa

(Received 28 April 2010; final version received 25 September 2010)

Unemployment and earnings inequality have moved together remarkably closely in South Africa in recent years. This article explores the relationship between unemployment and earnings inequality in South Africa, investigating the extent to which changes in unemployment can account for changes in earnings inequality. Static and dynamic decompositions of earnings inequality by employment status reveal the centrality of unemployment in accounting for both the level and trend of earnings inequality. The distribution of employment in the formal and informal sectors is found to be of lesser importance in explaining earnings inequality, as is wage dispersion within each of these categories. The findings point to the central importance of reducing unemployment in South Africa if the extremely high levels of inequality are to be reduced.

Keywords: inequality; earnings distribution; unemployment; labour market; South Africa

JEL Classifications: D30; E24; J31

1. Introduction

The combination of inequality and unemployment is almost uniquely high in South Africa when compared internationally. The levels and racialised character of both inequality and unemployment in South Africa are undoubtedly to a large extent a product of the country’s apartheid legacy. This article investigates particular aspects of earnings inequality in South Africa, focusing on the ways in which the rate of unemployment, the formal/informal composition of the employed, and wage dispersion amongst each of the formal and informal sectors contribute to earnings inequality. The aggregate relationship between unemployment rates and earnings inequality is investigated here. Furthermore, decomposition techniques are used to quantify the relative contributions of the unemployment rate and various other dimensions of labour force structure and the earnings distribution to overall earnings inequality. The methodology used here to explore these questions could also be useful for application to other countries.

The remainder of this introduction briefly considers some of the literature relevant to the analysis of this article. Section 2 looks at some salient aspects of earnings inequality in South Africa, and the trends in earnings inequality and unemployment. The empirical analysis of the relationship between particular aspects...
of the labour market and earnings inequality is contained in Section 3. Section 4
discusses the results and some possible policy implications, and concludes.

A prominent view in the mainstream literature is that there is a trade-off between
increasing earnings inequality and increasing unemployment. This has been consid-
ered to explain the differences in patterns of unemployment and income or wage
inequality when comparing the US and Europe, and to a lesser extent the US and
Canada (see for example Storer and Van Audenrode 1998, and Ayala et al. 2002).
Labour market characteristics (such as the degree of centralisation in bargaining sys-
tems, the system of unemployment benefits, and the level and coverage of a minimum
wage) are typically considered to have opposite effects on earnings inequality and on
unemployment. Accordingly, changes in these characteristics within a country would
be expected to affect earnings inequality and unemployment in opposite directions.

While the existing South African literature does not directly address the ques-
tions investigated here, several studies do find labour market factors to be important
in explaining inequality. Leibbrandt and Woolard (2001) find labour market activi-
ties to be significant contributors to households’ movements into and out of poverty
in the Kwazulu-Natal province. Bhorat et al. (2000) use the 1995 Income and
Expenditure Survey to decompose the Gini coefficient by income source, finding
wage and salary income to be the component that contributes most to overall
income inequality. Leite et al. (2006) study post-Apartheid earnings inequality in
South Africa and decompose earnings inequality along various lines, including
whether the person is an employee, self-employed, or both. They find that between-
group inequality according to these categories accounted for about 8.6% of inequal-
ity in 1997/1998, but this declines to zero or close to zero by 2004.

Of the studies that consider the relationships between labour markets and
inequality in South Africa, important linkages are found, even though older data is
used. The existing literature does not, however, directly address the issues investigat-
gated here, specifically in terms of the overall relationship between trends in unem-
ployment and in earnings inequality, as well as in terms of quantifying the relative
contributions of the unemployment rate, employment status, the formal/informal
structure of employment, earnings dispersion among the employed, and other rele-
vant features of the labour market to earnings inequality.

2. Inequality in South Africa

2.1. Earnings inequality in South Africa

Table 1 summarises the current level of earnings inequality among the employed in
South Africa, using alternative measures of inequality. Throughout the article,
‘earnings inequality’ refers to inequality in earnings on an individual basis. Earnings
inequality is extremely high: for instance, the Gini coefficient of earnings amongst
the employed is 0.60.

In recent years earnings grew (proportionally) most for those in about the lower
third of the distribution, albeit less for the very lowest end. Growth incidence
curves of earnings (amongst the employed) between 2001 and 2007 are shown in
Figure 1. Surprisingly, earnings appear to have fallen in real terms for middle-upper
earners. However, the top end of the distribution benefited from earnings growth
above that of the rest of the top half of the distribution. To the extent that there has
been some ‘redistribution’ towards the lowest earners, the relative losers have been
not the high income earners but the middle and upper-middle parts of the distribution.

2.2. Trends in inequality and unemployment

Both inequality and unemployment peaked in 2002 and have since declined (at least until the onset of the global economic downturn), albeit at a slow pace given their severity. There has been a close relationship between unemployment and earnings inequality, both among the labour force and among all ‘working age’ adults, as evident from Figure 2.3

These close relationships would be partially explained by the fact that higher unemployment means that a lower proportion of the labour force and of the working age adult population receive earnings and hence inequality would be higher in a straightforward ‘compositional’ sense. Figure 3 therefore shows the relationship between earnings inequality amongst the employed only and unemployment. These series exclude the direct compositional effect of unemployment on labour force or adult earnings inequality.

Interestingly, there is still a very clear positive relationship between unemployment and earnings inequality amongst the employed. This suggests that there is a
relationship beyond the ‘compositional’ channel. The relationship remains similarly close when other measures of inequality are used. In all cases the correlation coefficient between unemployment and earnings inequality amongst the employed is over 80% and is statistically significant.

It is remarkable how close unemployment and earnings inequality have moved together over time. It would not be especially surprising to find a positive relationship between unemployment and overall income inequality (where this includes the unemployed) given that unemployment would directly affect income. What is found here, however, is a positive relationship between earnings inequality amongst the employed and unemployment.

3. Decomposing the effects of unemployment and employment structure on earnings inequality

The effects of unemployment and of other selected dimensions of labour market structure on earnings inequality will now be investigated by decomposing earnings
inequality by population subgroups, where the subgroups are various categories of
the labour market. Data availability restricts this analysis to earnings inequality (and
not income inequality more broadly).

3.1. Static decomposition of earnings inequality

The use of decomposition analysis by population subgroups has been used interna-
tionally in the analysis of inequality by subgroups such as regions or racial groups.
The intuition behind the decomposition of inequality by subgroups is to divide a
population into discrete subgroups, with partitioning on the basis of distinct and
mutually exclusive personal or group characteristics, and to compute the inequality
within and between each of these subgroups. The ‘between-groups’ component is
calculated across the entire population and shows the differences in the mean of
earnings between groups. This basically indicates how much inequality there would
be, were there no inequality within each subgroup, i.e. if every member of the
group received the mean earnings of the group such that inequalities between
groups were the only source of inequality. The ‘within-groups’ inequality is a
weighted sum of the inequality within each subgroup, and shows how much
inequality there would be if there was no inequality between the groups. These two
components sum to total inequality.

Following Mookherjee and Shorrocks (1982), we define $\mu$ as the mean earnings
of the population (with population as defined in each decomposition below); and $y_i$
as the earnings of individual $i$ for $i = 1, 2, \ldots n$.

The population can be partitioned into subgroups based on labour market status
(as set out below for the various decompositions), with:

- $N_k$ the subset of individuals in subgroup $k$;
- $n_k$ members of subgroup $k$;
- $\mu_k$ as the mean earnings of subgroup $k$;
- $v_k = \frac{n_k}{n}$ the proportion of the population in labour market subgroup $k$; and
- $\lambda_k = \frac{\mu_k}{\mu}$ the subgroup mean earnings relative to the aggregate population mean.

Earnings inequality (measured by mean log deviation) is then decomposed as follows:

$$I_0 \frac{1}{n} \sum_i \ln \left( \frac{\mu}{y_i} \right) = \sum_k v_k I_0^{\text{within-groups component}} + \sum_k v_k \ln \left( \frac{1}{\lambda_k} \right)$$

In the initial analysis, the two subgroups are the employed and the unemployed.
The static decomposition of earnings inequality presented here indicates how much
of earnings inequality can be accounted for by the fact that the employed receive
earnings whereas the unemployed do not, and how much can be accounted for by
inequality in earnings amongst the employed. Given how the decompositions are
set up here, the within-groups component essentially measures the relative impor-
tance of inequality amongst the employed. The between-groups component basically
measures how much of earnings inequality is explained by the difference between
the mean earnings of those employed with the zero earnings$^4$ of those not working.
While both components would be known to be positive a priori, the decomposition analysis can shed light on the relative importance of the two components.

The results are shown in Table 2, for decompositions on three alternative populations. These populations are the labour force – each of the official and expended definitions – and the working age adult population (between the ages of 19 and 65 inclusive). For the two labour force populations, the two groups are the employed and the unemployed; and in the analysis of working age adults the two groups are those working and those not working (i.e. including both the unemployed and those outside of the labour force) between aged 19 and 65 inclusive.

For ease of interpretation, the between-groups figure is shown here as

\[
\frac{100 \sum_k v_k \ln \left( \frac{1}{\lambda_k} \right)}{I_0}
\]

indicating how much of earnings inequality is explained by inequality between the employed and the unemployed; while the within-groups figure

\[
\frac{100 \sum_k v_k I_k}{I_0}
\]

shows how much of total inequality is explained by inequality amongst the employed.

A key finding from this part of the analysis is the importance of between-group inequality in accounting for earnings inequality (among the labour force or among the working age population). As would be expected, the relative importance of between-group inequality rises as the population being analysed expands, since the proportion of non-earners within the sample increases. The contribution of earnings inequality within the employed to broader earnings inequality – shown here by the within-groups component – ranges between just 7% for the working-age adult population and 22% for the labour force (officially defined). This contribution is driven by very low earnings amongst some of the employed, and by the high degree of earnings inequality amongst the employed more broadly.

Next, those working are subdivided into two categories: those employed in the formal sector, and those employed in the informal sector (including domestic workers). This investigates the impact of not only the rate of unemployment but also of this aspect of employment structure on earnings inequality. To contextualise the decomposition analysis that follows, Table 3 compares earnings in the formal and informal sectors. Ninety-one percent of earnings go to people employed in the formal sector, with average earnings significantly lower in the informal sector.

Table 2. Static decomposition of current earnings inequality by employment status (%).

|                          | Between-groups | Within-groups | Total |
|--------------------------|----------------|---------------|-------|
| Labour force – official definition | 78.34          | 21.65         | 100   |
| Labour force – expanded definition | 87.02          | 12.98         | 100   |
| Working age adults       | 92.98          | 7.02          | 100   |

Note: Calculations based on data from September 2007 LFS.
Inequality of earnings is roughly similar between the formal and informal sectors when measured with the Gini, but differ more when using mean log deviation. The latter measure is more sensitive to the bottom end of the distribution, and this difference is probably picking up the fact that earnings tail off to lower levels at the bottom of the distribution in the informal sector as compared to in the formal sector.

Earnings inequality is decomposed according to labour market status, with the groups being the formally employed; the informally employed; and the unemployed or not working. The results are summarised in Table 4. The between-groups component accounts for the vast majority of earnings inequality – between 83% in the case of the officially defined labour force and 94% for all working age adults. Furthermore, this component is found to be more important here in accounting for overall earnings inequality than in the previous decompositions (where all the employed were treated as a single group). This is probably due to the higher level of disaggregation used here, and that average earnings are significantly higher in the formal than the informal sector.

The level of wage dispersion within each of the formal and informal sectors does contribute to overall earnings inequality. But of much greater importance are the gaps between the average earnings of the formal and informal sectors and between these and the zero earnings received by the unemployed. These findings might suggest that reducing the rate of unemployment, as well as (of somewhat lesser importance) closing the gap between formal and informal sector earnings or moving people from the informal to formal sectors, would be central to reducing the overall level of earnings inequality. Reducing earnings dispersion within each of the formal and informal sectors is of lower importance in this regard.

Table 3. Current earnings inequality in the labour force, by labour market status.

|                      | Formally employed | Informally employed |
|----------------------|-------------------|---------------------|
| Population share     | 56.3%             | 20.6%               |
| Earnings share       | 91.1%             | 8.9%                |
| Gini                 | 0.54              | 0.57                |
| Mean log deviation   | 0.57              | 1.35                |

Notes: Population is the labour force. Calculations based on data from September 2007 LFS.

Table 4. Static decomposition of earnings inequality by labour market category.

|                             | Between-groups | Within-groups | Total |
|-----------------------------|----------------|---------------|-------|
| Labour force – official definition | 82.55          | 17.44         | 100   |
| Labour force – expanded definition | 89.67          | 10.33         | 100   |
| Working age adults          | 94.09          | 5.91          | 100   |

Notes: The groups in this decomposition are the formally employed, informally employed, and the unemployed in the cases where the reference population is the labour force; and the formally employed, informally employed, and those not working where the reference population is the working age adult population. Calculations based on data from September 2007 LFS.
3.2. Dynamic decomposition of earnings inequality

Next, a dynamic decomposition methodology is used to analyse the changes in earnings inequality between 2001 and 2007, in order to explain how much of these changes can be accounted for by changes in the particular aspects of labour market structure. Specifically, the analysis seeks to identify how much of the changes in earnings inequality can be accounted for by changes in factors such as unemployment, earnings dispersion amongst the employed, and differences in earnings between the formal and informal sectors. The dynamic decomposition is based on the method pioneered by Mookherjee and Shorrocks (1982).

Further define: \( \bar{\nu}_k = \frac{1}{2} (v_k^t + v_k^{t+1}) \) for time periods \( t \) and \( t+1 \), and similarly for \( \bar{I}_0^k \) and \( \log \bar{\lambda}_k \), and \( \theta_k = v_k \lambda_k \), the income share of group \( k \). Then the change in inequality can be decomposed as follows:

\[
\Delta I_0 = \sum_k \bar{\nu}_k \Delta I_0^k + \sum_k \bar{I}_0^k \Delta \nu_k + \sum_k (\bar{\lambda}_k - \log \bar{\lambda}_k) \Delta \nu_k + \sum_k (\bar{\theta}_k - \bar{\nu}_k) \Delta \log \mu_k
\]

This analysis allows for identification of the relative contributions to a change in earnings inequality of changes in earnings dispersion within each of the subgroups, of changes in the relative proportions of each of the subgroups; and of changes in the relative income of the subgroups. As with the static decompositions of inequality set out in the previous section, this analysis begins with a simple decomposition of inequality between just two groups – the employed and the unemployed. In this case, since the total population in the initial decompositions is the labour force, the second component (changes in the shares of subgroups) measures the relative (direct) contribution of changes in the rate of unemployment.

The results are shown for two periods: the episode in which both unemployment and inequality were increasing (2001–2002) and the episode in which both were falling (2002–2007). Applying the decomposition to these two periods separately avoids the high degree of volatility between individual years or between the biannual surveys, while also picking up the potentially different dynamics of these two distinct periods rather than mixing them together. The results are shown in Tables 5 and 6 (using the official and expanded definitions of unemployment respectively), in percentage form. The components sum to 100% in the first period and \(-100\%\) in the second period, since inequality rose in the first period and fell in the second.

The most important result arising from this analysis is the importance of changes in the unemployment rate in explaining changes in earnings inequality within the labour force. During the first period, in which both unemployment and inequality rose, increases in the unemployment rate accounted for 77% of the increase in earnings inequality within the labour force. Both unemployment and
inequality fell in the second period, with the decrease in the unemployment rate explaining 72% of the decrease in inequality. These results highlight the huge importance of the unemployment rate in explaining earnings inequality, both during rises and falls of inequality and unemployment.

Inequality amongst the employed contributed to a relatively small extent (16%) to the increase in inequality amongst the entire labour force in the first period, and to a somewhat larger extent (40%) to the decrease in inequality in the second. It is interesting that the contribution of inequality amongst the employed moved in the same direction as trends in overall labour force inequality as well as in the unemployment rate in both periods (as evident from the sign on the component ‘effect of changes in earnings inequality amongst employed’). While decomposition analysis cannot discern causality between the components, it is noteworthy that the contributions of changes in the rate of unemployment and changes in earnings inequality amongst the employed moved in the same direction in both periods.

The third component of the decomposition is changes in between-group inequality. This captures the effect of the change in relative mean earnings of the employed and unemployed on changes in overall earnings inequality of the labour force, and is essentially a residual factor in this particular decomposition. This is the only component with the same sign (positive) in the two periods, indicating that it contributed to the rise in inequality in the first period and mitigated the fall in inequality in the second period. However, the contribution was relatively small.

The results are similar for the expanded definition of the labour force (including ‘discouraged’ job-seekers), as shown in Table 6. The main difference is that

|                  | Period 1     | Period 2     |
|------------------|--------------|--------------|
|                  | (2001-1–2002-1) | (2002-1–2007-2) |
| Effect of changes in earnings inequality amongst employed | 11.1          | −46.4        |
| Effect of changes in unemployment rate | 80.4          | −77.9        |
| Effect of changes in between-group inequality | 8.6           | 24.3         |
| Total            | 100          | −100         |

Note: Calculations based on LFS data (February/March 2001, March 2002, and September 2007).
changes in the unemployment rate account for an even higher proportion (around 80%) of total changes in earnings inequality, in both the periods of rising and falling unemployment. This underlines the centrality of unemployment, not only in accounting for the greater part of earnings inequality in a static sense, but also in accounting for most of changes in earnings inequality over time.

As with the static decomposition, this analysis is next extended by subdividing the employed into those working in the formal and informal sectors. The changes in labour force earnings inequality between 2001 and 2007 are decomposed according to three subgroups: the formally employed; the informally employed; and the unemployed.

These results, summarised in Tables 7 and 8, reinforce those from the dynamic decomposition into employed and unemployed. The most important factor explaining changes in earnings inequality is changes in labour force structure – referring here to the proportions of the labour force that are employed in the formal sector, employed in the informal sector, and unemployed, respectively. This component accounts for most of the changes in earnings inequality both during the period of rising inequality and during the subsequent period in which inequality fell. It was especially important during the latter period (accounting for 77% of the fall in earnings inequality when the official measure of the labour force is used, and 83% under the expanded definition). It can also be noted that earnings dispersion amongst the employed changed in the same direction as unemployment (also as overall earnings inequality) in both periods; the contribution to overall earning

Table 7. Dynamic decomposition of earnings inequality by formal employment/informal employment/unemployment [official definition].

| Period 1 (2001-1–2002-1) | Period 2 (2002-1–2007-2) |
|--------------------------|--------------------------|
| Effect of changes in earnings inequality amongst employed | 28.3 | −28.1 |
| Effect of changes in proportions formally employed / informally employed / unemployed | 58.1 | −77.4 |
| Effect of changes in between-group inequality | 13.6 | 5.5 |
| Total | 100.0 | −100.0 |

Note: Calculations based on LFS data (February/March 2001, March 2002, and September 2007).

Table 8. Dynamic decomposition of earnings inequality by formal employment/informal employment/unemployment [expanded definition].

| Period 1 (2001-1–2002-1) | Period 2 (2002-1–2007-2) |
|--------------------------|--------------------------|
| Effect of changes in earnings inequality amongst employed | 19.9 | −32.7 |
| Effect of changes in proportions formally employed / informally employed / unemployed | 69.1 | −82.9 |
| Effect of changes in between-group inequality | 11.0 | 15.6 |
| Total | 100.0 | −100.0 |

Note: Calculations based on LFS data (February/March 2001, March 2002, and September 2007).
inequality was, however, significantly weaker than was the contribution of changes in labour force structure.

4. Discussion
The empirical findings point strongly to the centrality of unemployment for understanding of inequality in South Africa. A surprisingly close positive relationship between the trends in unemployment and in earnings inequality amongst the employed over time can be observed, with a correlation coefficient of over 80% (which is robust to the use of various alternative measures of inequality). This suggests that – at least for the period analysed and for the ranges of inequality and unemployment during that period – there might not be a trade-off between inequality and unemployment.

The relevance of unemployment to inequality is underscored and quantified by the results from the static and dynamic decomposition analyses of earnings inequality. Even insofar as a positive contribution of unemployment to overall earnings inequality might have been expected, these results are useful for measuring the relative contributions of unemployment and of other factors. The rate of unemployment is found to account for the bulk of earnings inequality. Furthermore, changes in the unemployment rate account for most of the changes in inequality, both during the initial rise in inequality and during the subsequent decline. Changes in the unemployment rate turn out to be much more significant in accounting for changes in overall earnings inequality than do factors such as changes in the proportions of workers in the formal versus the informal sectors, or changes in wage dispersion within each of the formal and informal sectors.

Together with the observed positive co-movement between the unemployment rate and the degree of earnings inequality among the employed, these results could suggest that rather than there being a trade-off between reducing unemployment and reducing inequality, similar policies might be able to address both. While this article has not specifically investigated the effects of earnings dispersion on the employed on unemployment, at the least the evidence discussed here does not provide support for the notion that greater earnings dispersion would be more conducive to reducing the level of unemployment, or conversely that reducing earnings inequality would worsen unemployment. Rather, the results suggest that addressing the crisis of unemployment is absolutely central to reducing South Africa’s extremely high levels of inequality.

Earnings dispersion amongst the employed, and the proportions of people in the formal and informal sectors, are also important although lesser contributors to inequality among the labour force and among working age adults. Having established the centrality of addressing unemployment for the reduction of inequality, it also cannot be said that just ‘any jobs’, however badly paid, would really be a solution to the problem of high levels of inequality in South Africa. An increase in the dispersion of earnings amongst the employed, or an informalisation of employment, could erode the potential inequality-reducing effects of large-scale employment creation. A massive expansion of decent employment opportunities, particularly for the low-skilled and semi-skilled, could be the most important means of bringing down overall inequality in South Africa.

The effects on overall earnings inequality of the unemployed gaining employment would depend both directly on the earnings of those gaining employment and
indirectly on the effects of the reduced rate of unemployment and the new employment on the earnings distribution of those already employed. The unemployed are generally relatively low-skilled and hence it is probable that, were they to gain employment, they would thus swell the lower end of the earnings distribution. Ceteris paribus, this would tend to reduce inequality. The indirect effects on the earnings of those already employed are less predictable. It seems most likely that reduced unemployment would tend to most strongly affect the earnings of those with whom the (previously) unemployed have the most similar profiles, by putting upward pressure on their earnings. In part, this could be through ‘reserve army’-type effects of lower unemployment on the bargaining power of the employed. It thus seems credible that reduced unemployment could have indirect effects in pushing up the earnings of the lower part of the distribution of those already employed. This would also tend to reduce inequality (although the effects would be quite sensitive to which measure of inequality was used).

The most important dynamic underlying future distributional changes is likely to be through the labour market, in terms of both employment creation (or losses) and the distribution of earnings amongst the employed. It is improbable whether South Africa’s inequality could be brought down to ‘normal’ standards of inequality by international standards without increased demand for low- and semi-skilled labour, and to a lesser extent through a closing of wage gaps.

Acknowledgements
Useful comments from two anonymous referees and from the Editor of this journal are gratefully acknowledged.

Notes
1. All data used here are derived from the Labour Force Survey (LFS), a nationally representative biannual survey conducted by the country’s official statistics agency, Statistics South Africa. All original datasets were accessed through the South African Data Archive.
2. Comparable data are not available for earlier years.
3. These relationships also hold when the ‘expanded’ measure of unemployment is used. The difference between the official and expanded definitions of unemployment is that the former excludes from the labour force people who have not looked for work or taken steps to start a business in the four weeks prior to the survey interview. Both measures are limited to people aged between 15 and 65 who did not have a job or business in the seven days prior to the interview and were available to take up work within two weeks of the interview.
4. Actually, the imputed earnings of R0.01 per month for computational purposes.
5. The categorisation of the formal and informal sectors used in this analysis is based on the definitions used by Statistics South Africa, where LFS respondents are allocated based on their own response as to whether their employer is in the formal or informal sector.
6. That is, the effect of changes in earnings inequality among the employed is shown as $100 \frac{\sum \tilde{v}_k \Delta l_0^k}{\Delta l_0}$; the effect of changes in the rate of unemployment is shown as $100 \frac{\sum \tilde{l}_0^k \Delta v_k + \sum (\tilde{\bar{v}}_k - \log \bar{z}_k) \Delta v_k}{\Delta l_0}$; and the effect of changes in between-group inequality (i.e. the relative income of subgroups) as $100 \frac{\sum (\tilde{\bar{v}}_k - \bar{v}_k) \Delta \log \mu_k}{\Delta l_0}$. 

F. Tregenna
References
Ayala, L., R. Martinez, and J. Ruiz-Huerta. 2002. Institutional determinants of the unemployment-earnings inequality trade-off. *Applied Economics* 34: 179–95.
Bhorat, H., M. Leibbrandt, and I. Woolard. 2000. Understanding contemporary household inequality in South Africa. *Studies in Economics and Econometrics* 243: 31–52.
Leibbrandt, M., and I. Woolard. 2001. The labour market and household income inequality in South Africa: existing evidence and new panel data. *Journal of International Development* 13: 671–89.
Leite, P.G., T. McKinley, and R.G. Osorio. 2006. The post-apartheid evolution of earnings inequality in South Africa, 1995–2004. International Poverty Centre Working Paper No. 32, October 1996.
Mookherjee, D. and A. Shorrocks. 1982. A decomposition analysis of the trend in UK income inequality. *The Economic Journal* 92: 886–902.
Storer, P.A., and M.A. van Audenrode. 1998. Exploring the links between wage inequality and unemployment: A comparison of Canada and the US. *Canadian Public Policy/Analyse de Politiques* 241: S233–53.
Appendix. Processing of LFS data

The empirical analysis was undertaken using the full datasets of the LFS, February 2001–September 2007. Certain key elements of the processing of the original data are summarised below.

**Screening of high incomes**

A small number of original observations were excluded on the grounds of their unrealistically high reported earnings, particularly in the light of the occupations and other personal characteristics of the respondents.

**Treatment of earnings reported in brackets**

Respondents unwilling or unable to state their actual earnings could instead indicate which of 14 brackets their income falls within. This poses a problem for computations requiring income as a continuous variable. The mean income of people who reported actual incomes were calculated, by bracket, for each year. These were then assigned to the people in the same bracket who only identified a bracket.

**Treatment of non-responses on earnings**

There is a problem of non-response to the earnings question in the LFS, where some respondents refuse to disclose their earnings or indicate that they do not know their earnings, or no data are entered for other reasons. Since it is likely that the earnings variable is missing not at random, dropping of these observations could lead to bias in empirical analysis based on this variable. This problem was addressed by imputing earnings for missing observations. The method used was hotdeck imputation, in which a vector of respondent characteristics relevant to earnings was used to impute missing earnings, based on the characteristics of individual nonrespondents and other individuals with similar characteristics who did disclose their earnings.