UNIVERSITY OF CAPE TOWN

Dissecting post-apartheid labour market developments: Decomposing a discrete choice model while dealing with unobservables

Martin Wittenberg[1]

Working Paper Number 46

[1] School of Economics, University of Cape Town

Dissecting post-apartheid labour market developments: Decomposing a discrete choice model while dealing with unobservables*

Martin Wittenberg
School of Economics
University of Cape Town

January 2007

Abstract

The abolition of apartheid should have improved the employment prospects of black South Africans. The reality seems to have been different, with rising unemployment rates. Disentangling the real trends from changes in measurement and sampling design has proved to be difficult. We tackle this issue by means of an new methodology for decomposing changes in a proportion.

Our approach is based on a methodology presented by Lemieux for continuous variables. In particular we show how we can construct counterfactual data at the individual level controlling for unobservable effects. We show that this methodology has many attractive features when compared to other approaches available. In particular it lends itself to graphical analyses.

We use this methodology to explore changes in the proportion of African men being employed, unemployed and not economically active in South Africa in the post-apartheid period. Our results suggest that changes in the characteristics of these men have made them more employable over time, but that the propensity to be employed has declined. One might say that the human and social capital of these men has improved, but that the returns on that capital have declined. The net effect has been to leave measured employment more or less static. Changes in their characteristics and in their propensity to be economically active have both worked towards increasing the participation rate. As a consequence unemployment has risen over time.

The analysis confirms that there are important measurement changes between different national surveys.

Keywords: decomposition, discrete choice models, South Africa, employment, unemployment, participation

JEL Codes: C25 J21

*I would like to thank Taryn Dinkelman and Thomas Kniesner for insightful comments and Nicola Branson for sterling research assistance.

1 Introduction

South African social engineering has been on such a scale that it might be thought of as a social laboratory in which many huge "unnatural" experiments have been carried out simultaneously. For a social scientist this presents many opportunities, but also many challenges. How does one begin to pick apart the impact of different policies? The labour market is one arena which saw many distortions. Black South Africans were subject to the "colour bar" which excluded them from certain occupations. Controls on migration kept many of them in the rural areas and segregation of the schooling system made the accumulation of skills more difficult. The deracialisation of the South African economy after 1994 should therefore have led to big shifts in the employment of black South Africans. The reality seems to have been quite different with increasing rates of unemployment during the first decade of democracy (Banerjee, Galiani, Levinsohn and Woolard 2006). Understanding the dynamics of these changes is bedevilled by measurement problems. In principle it should be easy to track the shifts. Since 1993 there have been annual (or even biannual) national household level surveys that have attempted to measure employment and labour force participation. Regrettably, however, changes in sampling design have made attempts to compare trends over time very difficult (Branson and Wittenberg 2006).

Our approach in this paper is to analyse the changes by means of a Lemieux-style decomposition (Lemieux 2002) for a discrete choice model. This, to our knowledge, is the first time that this has been done. This decomposition allows us to look at the changes in two ways. We ask what the South African labour market would have looked like if the individuals sampled in previous years, had faced the labour market conditions (coefficients) of March 2004. We also ask what the labour market would have looked like if the individuals sampled in March 2004 had faced the conditions of previous years, while keeping their characteristics, including any unobservable traits that impacted on their labour market status in 2004. These should bound the actual changes even with the shifts in sampling design.

The decomposition reveals that changes in the average characteristics of African males[1] made them more employable over time. This is as we would have expected. Changes in education and freer migration since the end of apartheid should have improved the job market prospects of these individuals. This effect is offset, however, by a declining overall propensity to be employed. One might say that while the human and social capital of African males has improved over time, the returns attached to that capital have declined. This might be due to changes in the position of South Africa in the global economy. The net effect of these shifts in characteristics and returns is to leave the overall proportion of the population employed fairly constant. Our analysis reveals also a major participation "shock", driven both by changes in characteristics and in the underlying propensity to be active. This is reflected in a large increase in the proportion unemployed.

Our analysis confirms that different surveys pick up markedly different levels of employment and participation. In particular, the 1995 survey which has been used to benchmark many discussions of post-apartheid trends seems to over-capture employment. This will obviously affect the inferences drawn about labour market trends since the advent of democracy. By "standardising" the different data series through our decomposition we pinpoint a number of

[1] It is impossible to avoid racial terminology when discussing South Africa. We use the term "African" to refer to black South Africans who were not classified "coloured" or "Indian" under the apartheid regime.

other anomalous data sets. Despite the fact that these surveys were run by the same organisation and to similar specifications, these fluctuations are out of line with purely random noise. Changes in sampling design, field work instructions, field work quality and coding may all have a role to play. This article should therefore be of interest to people interested in survey measurement issues as well as anyone concerned with the substantive questions about what happened in the aftermath of apartheid.

Indeed our decomposition technique should be of interest in itself, since many of the more interesting problems confronting the applied researcher involve counterfactuals: what would a particular woman earn if she were treated exactly like a man? What would the income distribution have been in year X if the conditions had been as in year Y? For many years the standard tool for these sorts of problems has been the Oaxaca-Blinder decomposition (Oaxaca 1973, Blinder 1973). The usual idea is to take the coefficients from a regression model estimated over one group or year and apply them to the other. Over time researchers started to understand that one of the limitations of this procedure was that it did not always deal satisfactorily with the unobservables: those variables that matter but that we cannot control for adequately in our models; or the errors that arise due to some other misspecification of the model. So, for instance, when we ask what wage a particular woman would earn, we should not simply assign the average wage of the corresponding men: we should take into account whether she seems to be earning above or below the level that we would have expected given her characteristics. Consequently some authors started to pay more explicit attention to the importance of the residuals (e.g. Juhn, Murphy and Pierce 1993). Recently several authors have extended these approaches and explored ways of decomposing entire distributions (diNardo, Fortin and Lemieux 1996, Melly 2005). Lemieux (2002), for instance, has shown how to track changes in the distribution of a variable, while keeping a set of explanatory variables constant, by means of a simple reweighting procedure. In principle this is easy to do and it lends itself to simple graphical analyses of the changes. In this paper we extend this Lemieux procedure to a discrete choice model.

Oaxaca-Blinder style decompositions of discrete choice models have been discussed in the literature for some time (Even and Macpherson 1990, Nielsen 1998, Yun 2004, Fairlie 2005). The fundamental approach is to model the propensity to be employed in year X by means of a logit or probit (or perhaps even multinomial logit) model and to impose the coefficients from year X onto the observations from year Y. In the process, however, no attention has thus far been paid to whether the individuals that we observe in year Y have been revealed to be more or less employable than those in year X. Our approach will be fairly simple. We know whether the individual that we are observing has, in fact, been measured to be employed. This imposes some constraints on the impact that the unobservables can have. When we create the counterfactual, i.e. when we consider whether this individual would have been employed in a different year, we take this additional information into account. Simply ignoring the unobservables seems to be problematic for at least two reasons. Firstly, it seems unlikely that the observables will ever be able to capture all of the determinants of employability. Secondly, it seems plausible that the model may be misspecified. The stringent assumptions underpinning the logit, probit or multinomial logit models are unlikely to be fully satisfied in our example. We will rely on the fact that these maximum likelihood models are reasonable approximations to the underlying "true" model (White 1982). By correcting our models for the residuals we hope that our inferences will be even more robust. We will show that correcting for the unobservables can empirically make a difference, particularly if we want to extend the analysis to subpopulations.

The plan of our discussion is as follows. In the next section we discuss the literature dealing with post-apartheid labour market trends. We show that there are some difficulties in comparing different data sets as they stand. Indeed there has been a vociferous debate about the reliability of some of the data sets. We then present our new decomposition method. We start (in section 3.1) by rehearsing the Lemieux decomposition for continuous random variables. In section 3.2 we show how this idea can be applied to a binary choice model. We note that this approach can be extended to a multinomial logit model, but relegate the technical details to an appendix (appendix A). We then return to our South African example. We first compare our version of the decomposition to the traditional version in section 4.1. We note that whether one uses a logit or multinomial logit model for the decomposition matters less than correcting for unobservables. The unobservables do make a difference! We then look at what the decompositions suggest by age group (in 4.2). Finally we present our analysis of the trends for the period 1993 to 2004 in section 4.3. We conclude with reflections on what the decomposition reveals and discuss the utility and limitations of the technique.

2 Employment, nonparticipation and unemployment in post-apartheid South Africa

South Africa began its first decade of democracy with high hopes. The lifting of all previous restrictions, active interventions in education, and affirmative action policies should have significantly improved the labour market position of black South Africans. At the end of the decade, however, the failure of "job creation" was seen as one of the most pressing policy issues (PCAS 2003, p.94). More depressingly, some analysts argued that the high unemployment level reached at that stage was, in fact, an "equilibrium" one (Banerjee et al. 2006). Others contended that the reason for this failure was due to the democratic state's protection of organised labour against the unemployed (Seekings and Nattrass 2006). The suggestion that a large chunk of black South Africans may have been the victims of democracy has been hotly contested. From early on there have been voices that have challenged the reliability of the data on which these analyses have been based.

Central to these debates are the data derived from a series of national household surveys for which the unit records are publicly available. The first of these was the 1993 Project for Statistics on Living Standards and Development (PSLSD) conducted under the auspices of the Southern African Labour and Development Research Unit (SALDRU) at the University of Cape Town. It was modelled on the Living Standards Measurement Surveys of the World Bank. It was a survey of 45,000 individuals in about 8,800 households. Since 1994 South Africa's official statistical agency, Statistics South Africa, has also conducted annual nationally representative household surveys in October. These "October Household Surveys" were more narrowly focussed but had larger sample sizes, generally around 30,000 households, but dropping to as low as 16,000 in 1996. The October Household Surveys in turn were discontinued after 1999 and replaced by biannual Labour Force Surveys.

A proper discussion of the changes in sampling design and sample sizes of these surveys would take us too far afield. Suffice it to say that even the surveys conducted by Statistics South

Africa show considerable variation. The "odd" years 1995, 1997 and 1999 were most similar in size and design and have consequently featured prominently in most analyses of post-apartheid employment trends. There are differences even here. The 1999 survey shows a considerable drop in the average household size over the earlier surveys. This trend continued in the subsequent Labour Force Surveys. Indeed some part of this change is driven by a very rapid increase in the number of one person households – so much so that the credibility of the data has been called into question (Wittenberg and Collinson forthcoming). Of course with the appropriate weights one could control for these changes. Unfortunately the weights released with the data sets are post-stratified to gross up to the population aggregates that Statistics South Africa's demographic model churns out for the period in question. As new census information becomes available, the weights change discontinuously. For instance, Casale, Muller and Posel (2004, p.984–5) note that such changes in weights probably account for an increase of 550 000 workers between September 2002 and March 2003. As it stands there are two breaks in the series. The earlier surveys (1993 and 1994) have weights based on the 1991 census. The surveys from 1995 through to 2002 have weights based on the 1996 census, while the more recent Labour Force Surveys are based on the 2001 census.

There have also been changes in the survey instruments, as discussed by Casale et al. (2004). They note, for instance, how the Labour Force Surveys devote considerable more attention to picking up informal activities. Indeed there is considerable evidence that the introduction of the Labour Force Surveys in February 2000 led to an upward revision in the estimates of employment, simply due to better recording of such activities (Casale and Posel 2002, pp.170–1). Similarly Klasen and Woolard (1999, 2000b) discuss quite carefully some of the differences in definitions used in different surveys. Nevertheless even when they correct for such differences, it does not shift the overall unemployment rate by more than the odd percentage point.

The main purpose of the Klasen and Woolard papers was to take issue with the International Labour Organization's (ILO's) country report on South Africa (Standing, Sender and Weeks 1996) which initially set the terms for the post-apartheid debate about levels of employment and unemployment. This report had criticised both the time series and household survey based estimates. The household level data (chiefly the PSLSD and the 1994 OHS) were criticised on the grounds of coverage and the definitions of unemployment. Klasen and Woolard corrected the household survey based estimates for many of the deficiencies highlighted, but concluded that these corrections would not alter the overall levels markedly. Indeed, they made the point that it was surprising in how many respects the picture presented by the three cross-sectional surveys available to them, viz. the PSLSD data set (for 1993), the October 1994 Household Survey and the October 1995 Household Survey, were congruent. Not only were the estimates of unemployment in broadly similar ranges, but the patterns of unemployment, particular in its racial and locational breakdown were very similar. This was despite the fact that each of the surveys was run according to a different methodology and the definitions of unemployment used were somewhat different.

The Klasen and Woolard paper led to an emerging consensus among South African labour economists that household survey evidence could and should be used to analyse the performance of the post-apartheid labour market (see also Bhorat 1999, Wittenberg 2002). For a while the debates moved onto considerations of what the appropriate definition of unemployment should be – the broad or the narrow definition (Kingdon and Knight 2004b, Kingdon and Knight 2006), how the unemployed differed from those not searching (Dinkelman and Pirouz 2002), and what the relationship between household structure and unemployment was (Klasen and Woolard 2000a).

The debate about the reliability of the household survey data resurfaced however, the instant that labour economists started to analyse changes in the level of employment and unemployment over time (Bhorat 2003, Bhorat 2004, Casale et al. 2004). The key question was whether the post-apartheid economy had succeeded in creating jobs for those deprived by the previous regime or whether changes in the economy (such as increased liberalisation) or in labour market regulations had led to "jobless growth" or even job losses. Indeed this issue became an important political one for the government since it found it difficult to accept that South Africa might have a large unemployment problem (Mbeki 2005).

Given the difficulties in comparing different data sets outlined here, our approach will be somewhat different. We will use a decomposition technique in order to standardise the comparisons (for a different application of a decomposition see Kingdon and Knight 2004a). The change in the proportion employed, unemployed or not participating between different surveys will be decomposed into:

- The change in the propensity (i.e. coefficients) while keeping the characteristics constant

  This change will encompass firstly changes in the underlying economic conditions which make it easier or more difficult to find employment; and secondly changes in the measurement and coding process, which incorporates how the questions were asked, what probing happened and how the answers were categorised.

- The change in characteristics while keeping the propensity for employment/unemployment/participation constant

  This change will encompass firstly, real changes in the underlying characteristics of the population, such as improved education and better location; and secondly changes in the sampled population which are not appropriately accounted for by the sampling weights.

- Residual changes unaccounted for by the coefficients or characteristics

It should be stressed that the decomposition analysis does not substitute for attempts to obtain sampling weights that would make the different surveys truly comparable or attempts to harmonise the categories and measurements from the different data sets. Nevertheless by separating out these effects, it does give us some bounds on what the impact of measurement changes or sampling changes might be. We now turn to present the decomposition itself.

3 Decomposing changes in a proportion

3.1 The Oaxaca-Blinder decomposition as extended by Lemieux

We begin our analysis by rehearsing the basics of the general decomposition of a linear model as presented by Lemieux (2002). The fundamental Oaxaca-Blinder model (Oaxaca 1973, Blinder 1973) assumes that the variable of interest can be written in linear form as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta}_t + u_{it} \tag{1}$$

The subscript i refers to the individual and t the group, in this case the time period; $\mathbf{x}_{it}$ is a $1 \times k$ vector of covariates, including a constant; $\boldsymbol{\beta}_t$ is a $k \times 1$ vector of fixed parameters and u_{it} is assumed to a be mean zero error process. Given the OLS estimates $\mathbf{b}_t$ of $\boldsymbol{\beta}_t$ the sample means will obey the relationship

$$\overline{y}_t = \overline{\mathbf{x}}_t \mathbf{b}_t \tag{2}$$

Consequently the change in sample means from period t to period s can be decomposed as

$$\overline{y}_t - \overline{y}_s = \overline{\mathbf{x}}_t \left(\mathbf{b}_t - \mathbf{b}_s\right) + \left(\overline{\mathbf{x}}_t - \overline{\mathbf{x}}_s\right) \mathbf{b}_s \tag{3}$$

where the first term shows the impact of changing the coefficients and the second term the impact of changes in the covariates. In the typical application (the analysis of log wages) the first effect can be thought of as a rate of return or price effect (often identified with discrimination) and the second the impact of differences in the characteristics (some times referred to as an "endowment" effect). Following Lemieux, let

$$\overline{y}_t^a = \overline{\mathbf{x}}_t \mathbf{b}_s \tag{4}$$

then we can write this decomposition as

$$\overline{y}_t - \overline{y}_s = \left(\overline{y}_t - \overline{y}_t^a\right) + \left(\overline{y}_t^a - \overline{y}_s\right)$$

The term $\overline{y}_t^a$ represents the hypothetical mean that would be observed in period t if the coefficients had been $\mathbf{b}_s$. For instance, if y_{it} represents log wages, then $\overline{y}_t^a$ might represent the average wages that black people (or women) might obtain if their characteristics were remunerated at the same rate as those of whites (or men). As Lemieux points out (following Juhn, Murphy and Pierce (1993)), we can rewrite this in terms of individual level hypothetical values. Individual observations can be written as

$$y_{it} = \mathbf{x}_{it} \mathbf{b}_t + \widehat{u}_{it} \tag{5}$$

where the OLS residuals $\widehat{u}_{it}$ are by construction orthogonal to the regression estimates and have mean zero. Let the individual level prediction be

$$y_{it}^a = \mathbf{x}_{it} \mathbf{b}_s + \widehat{u}_{it} \tag{6}$$

It follows that $\overline{y}_t^a = \overline{y_{it}^a}$ so all the calculations can be done on the individual level values, i.e. we can write the decomposition as

$$\overline{y}_t - \overline{y}_s = \left(\overline{y}_t - \overline{y_t^a}\right) + \left(\overline{y_t^a} - \overline{y}_s\right) \tag{7}$$

Working with the individual level imputations has an additional advantage, since Lemieux shows that a simple reweighting procedure can provide us with a way of allowing the characteristics to change, while keeping the coefficients (prices) constant. To fix the intuition in this regard, let us assume that our observations come in just two types: high education (type H) and low education (L). In this case we can assume that the covariate vector $\mathbf{x}$ consists of just two indicator variables, one for type H and one for type L. Consequently $\overline{\mathbf{x}}_t = \left[\begin{array}{cc} \theta_{Ht} & \theta_{Lt} \end{array}\right]$

where θ_{jt} is the proportion of the sample in period t that is type j. It is easy to verify that the corresponding regression coefficient b_{jt} is just $\overline{y}_{jt}$ i.e. we get the trivial decomposition

$$\begin{aligned}\overline{y}_t &= \overline{\mathbf{x}}_t \mathbf{b}_t \\ &= \theta_{Ht}\overline{y}_{Ht} + \theta_{Lt}\overline{y}_{Lt}\end{aligned}$$

We can now consider what would happen if we changed the distribution to that of period s, while keeping the coefficients (in this case the mean value within each group) constant, i.e. we are interested in the hypothetical construct

$$\overline{y}_t^c = \theta_{Hs}\overline{y}_{Ht} + \theta_{Ls}\overline{y}_{Lt}$$

Note that in this particular case $\overline{y}_t^c = \overline{\mathbf{x}}_s \mathbf{b}_t = \overline{y}_s^a$. Trivially we have

$$\theta_{Hs}\overline{y}_{Ht} + \theta_{Ls}\overline{y}_{Lt} = \psi_H \theta_{Ht}\overline{y}_{Ht} + \psi_L \theta_{Lt}\overline{y}_{Lt}$$

where $\psi_j = \theta_{js}/\theta_{jt}$, so we can calculate $\overline{y}_t^c$ from the individual level data using the individual level "weights" ψ_{it} where $\psi_{it} = x_{iH}\theta_{Hs}/\theta_{Ht} + x_{iL}\theta_{Ls}/\theta_{Lt}$, i.e. $\psi_{it} = \psi_j$ if individual i is of type j, i.e.

$$\overline{y}_t^c = \frac{1}{N}\sum_i \psi_{it} y_{it}$$

Lemieux adapts a procedure introduced by diNardo et al. (1996) to show that this reweighting procedure can be used even when the covariates are continuous. In this case the approach is to pool the samples of period t and s and to estimate the conditional probability $P(\tau = t|\mathbf{x}_{i\tau})$. This can be done by any appropriate binary choice model (e.g. probit or logit). The reweighting factor in this case is

$$\psi_{it} = \frac{P(\tau = s|\mathbf{x}_{i\tau})}{P(\tau = t|\mathbf{x}_{i\tau})}\frac{P(\tau = t)}{P(\tau = s)} \tag{8}$$

where $P(\tau = t) = 1 - P(\tau = s)$ is the unconditional probability that an observation belongs to period t. This is just the proportion of the joint sample that is from period t. Note that ψ_{it} will be equal to the previous definition in the case where $\mathbf{x}_{it}$ is given by the two indicator variables considered previously.

All of these results go through if the data come from a sample design where the inclusion probabilities are not equal. In this case

$$\overline{y}_t^c = \sum_i \omega_{it}\psi_{it} y_{it}$$

where ω_{it} are the appropriate sampling weights. If we define

$$\omega_{it}^a = \omega_{it}\psi_{it}$$

we get the results set out in table 1.

Given the linear nature of this model it is obvious that $\overline{y}_t^{ac} = \overline{y}_s$. Indeed in this case there is no advantage to be had from calculating the individually imputed values y_{it}^a or the weights ω_{it}^a. The power of the procedure is that it can be used with higher moments of the distribution or with non-linear functions of the variables. If we want to fix the covariates, but allow the coefficients to change, we use the values y_{it}^a. If by contrast we want to vary the characteristics while keeping the "prices" fixed, we use the weights ω_{it}^a with the original y_{it} values.

$\overline{y}_t = \sum_i \omega_{it} y_{it}$	The sample mean in period t
$\overline{y}_t^a = \sum_i \omega_{it} y_{it}^a$	The hypothetical mean that we would observe in period t if we were to fix the characteristics but allow the coefficients to change
$\overline{y}_t^c = \sum_i \omega_{it}^a y_{it}$	The hypothetical mean that we would observe in period t if we were to fix the coefficients but allow the covariates to change
$\overline{y}_t^{ac} = \sum_i \omega_{it}^a y_{it}^a$	The hypothetical mean that we would observe in period t if we were to allow the covariates and the coefficients to change

Table 1: Lemieux's decomposition of changes in a continuous dependent variable

3.2 Decomposing a binary choice model

A number of authors have discussed ways of extending the Oaxaca-Blinder methodology to binary choice models (Even and Macpherson 1990, Nielsen 1998, Yun 2004, Borooah and Iyer 2005, Fairlie 2005). The standard departure point is to start with a formulation of a binary choice model

$$p_{it} = F\left(\mathbf{x}_{it}\boldsymbol{\beta}_t\right) \tag{9}$$

which we can estimate by standard maximum likelihood techniques, i.e.

$$\widehat{p}_{it} = F\left(\mathbf{x}_{it}\mathbf{b}_t\right) \tag{10}$$

In the case of a logit model (as noted by Nielsen (1998)) we can use the fact that $\overline{p}_t = \overline{\widehat{p}}_t$, i.e. the sample proportion is always numerically equal to the average of the predicted probabilities. In the case of a probit this will hold approximately. We can then decompose the change in proportions as

$$\begin{aligned} \overline{p}_t - \overline{p}_s &\simeq \overline{\widehat{p}}_t - \overline{\widehat{p}}_s \\ &= \frac{1}{N_t}\sum_i \left(F\left(\mathbf{x}_{it}\mathbf{b}_t\right) - F\left(\mathbf{x}_{it}\mathbf{b}_s\right)\right) + \left\{\frac{1}{N_t}\sum_i F\left(\mathbf{x}_{it}\mathbf{b}_s\right) - \frac{1}{N_s}\sum_i F\left(\mathbf{x}_{is}\mathbf{b}_s\right)\right\} \\ &= \left(\overline{\widehat{p}}_t - \overline{\widehat{p}}_t^o\right) + \left(\overline{\widehat{p}}_t^o - \overline{\widehat{p}}_s\right) \end{aligned} \tag{11}$$

where

$$\widehat{p}_{it}^o = F\left(\mathbf{x}_{it}\mathbf{b}_s\right) \tag{12}$$

The first term of this decomposition is the "coefficient effect". It is a measure of how much the empirical proportion would change if the individual attributes were rewarded at the rate $\mathbf{b}_s$ rather than $\mathbf{b}_t$. The second term is the "covariates effect". It measures how much of a change we would expect if the coefficients were fixed and only the characteristics were allowed to change.

Note that perforce this decomposition relies on individual level imputations and counterfactuals. Nevertheless this procedure is not analogous to the one outlined in the previous section, since we would have effectively ignored the residual term used in the OLS imputations, e.g. in equation 6. It is clear why we wouldn't want to ignore these residuals in the OLS case: even with identical $\mathbf{b}$s, the hypothetical distribution would differ from the original distribution. If we are interested in the counterfactual distribution in its own right, or if we are interested in any statistics other than the mean, the implicit reduction in variance would lead to distorted results.

Can we adapt the Lemieux procedure for discrete choice problems? We can see the immediate problem if we try to apply the procedure to a linear probability model. We can write this model in the form of equation 1, estimate it by OLS and the resulting estimates will obey all of the decompositions as outlined in equations 2–5. One important point about equation 5, however, is that the residual will be able to take on only two values: $\widehat{u}_{it}$ can be $1 - \mathbf{x}_{it}\mathbf{b}_t$ or $-\mathbf{x}_{it}\mathbf{b}_t$. This means immediately that equation 6 does not produce valid data in the context of the LPM.

Our approach will be slightly different. We will start with the latent variable formulation of the binary choice model, i.e. we assume that there is a latent variable y_{it}^* defined as

$$y_{it}^* = \mathbf{x}_{it}\boldsymbol{\beta}_t + u_{it} \quad (13a)$$
$$y_{it} = \mathbf{1}\left(y_i^* > 0\right) \quad (13b)$$

where $\mathbf{1}$ is the indicator function. If we assume that u_{it} has cdf F, then

$$\begin{aligned} P\left(y_{it} = 1|\mathbf{x}_{it}\right) &= P\left(\mathbf{x}_{it}\boldsymbol{\beta}_t + u_{it} > 0\right) \\ &= 1 - F\left(-\mathbf{x}_{it}\boldsymbol{\beta}_t\right) \end{aligned}$$

Assuming that F is a distribution symmetric about zero we get the standard formulation

$$P\left(y_{it} = 1|\mathbf{x}_{it}\right) = F\left(\mathbf{x}_{it}\boldsymbol{\beta}_t\right)$$

With a suitable choice for F this can easily be estimated by maximum likelihood. Let the estimates be $\mathbf{b}_t$. The corresponding fitted value for the index is just $\mathbf{x}_{it}\mathbf{b}_t$. Implicitly these define a residual, i.e.

$$\widehat{u}_{it} = y_{it}^* - \mathbf{x}_{it}\mathbf{b}_t \quad (14)$$

Of course y_{it}^* is not observed, so neither is the residual. Nevertheless we can say something more specific about the distribution of $\widehat{u}_{it}$ given that $\widehat{u}_{it}$ has, asymptotically, the same distribution as u_{it}. If $y_{it} = 1$, we know that $y_i^* > 0$, i.e. $\widehat{u}_{it} > -\mathbf{x}_{it}\mathbf{b}_t$. If $y_{it} = 0$ we have $\widehat{u}_{it} \leq -\mathbf{x}_{it}\mathbf{b}_t$. Consequently $\widehat{u}_{it}$ will be distributed with F truncated at $-\mathbf{x}_{it}\mathbf{b}_t$.

Corresponding to equation 6 we will define

$$y_{it}^{*a} = \mathbf{x}_{it}\mathbf{b}_s + \widehat{u}_{it} \quad (15a)$$
$$y_{it}^a = \mathbf{1}\left(y_{it}^{*a} > 0\right) \quad (15b)$$

Since $\widehat{u}_{it}$ is not observed, neither is y_{it}^{*a} and consequently we will not always be able to deduce what y_{it}^a is either. Nevertheless we can be quite specific about the probability

$$\begin{aligned} p_{it}^a &= P\left(y_{it}^a = 1|\mathbf{x}_{it}\mathbf{b}_s, \widehat{u}_{it}\right) \\ &= P\left(\mathbf{x}_{it}\mathbf{b}_s + \widehat{u}_{it} > 0\right) \\ &= P\left(\widehat{u}_{it} > -\mathbf{x}_{it}\mathbf{b}_s\right) \end{aligned} \quad (16)$$

	$y_{it}=1$	$y_{it}=0$
$\mathbf{x}_{it}\mathbf{b}_s > \mathbf{x}_{it}\mathbf{b}_t$	$p^a_{it}=1$	$p^a_{it}=\frac{F(\mathbf{x}_{it}\mathbf{b}_s)-F(\mathbf{x}_{it}\mathbf{b}_t)}{1-F(\mathbf{x}_{it}\mathbf{b}_t)}=\frac{\widehat{p}^o_{it}-\widehat{p}_{it}}{1-\widehat{p}_{it}}$
$\mathbf{x}_{it}\mathbf{b}_s < \mathbf{x}_{it}\mathbf{b}_t$	$p^a_{it}=\frac{F(\mathbf{x}_{it}\mathbf{b}_s)}{F(\mathbf{x}_{it}\mathbf{b}_t)}=\frac{\widehat{p}^o_{it}}{\widehat{p}_{it}}$	$p^a_{it}=0$

Table 2: Values of p^a_{it}

Since we know the distribution of $\widehat{u}_{it}$ and the values of $\mathbf{x}_{it}\mathbf{b}_s$ and $\mathbf{x}_{it}\mathbf{b}_t$ this probability can be easily computed. Indeed if $\mathbf{b}_s=\mathbf{b}_t$ then $p^a_{it}=0$ if $y_{it}=0$ and $p^a_{it}=1$ if $y_{it}=1$. In short $p^a_{it}=y_{it}$ for every observation, so p^a_{it} seems the obvious proxy for y^a_{it}. For the cases where $\mathbf{b}_s\neq\mathbf{b}_t$ we can summarise the values of p^a_{it} in table 2.

These expressions can be derived quite easily by Bayes' Law if we observe that they are just conditional probabilities, e.g.

$$\begin{aligned} P\left(y^a_{it}=1|\mathbf{x}_{it}\mathbf{b}_s,\widehat{u}_{it}\right) &= P\left(y^a_{it}=1|\mathbf{x}_{it}\mathbf{b}_s,y_{it}=1\right) \\ &= \frac{P\left(\widehat{u}_{it}>-\mathbf{x}_{it}\mathbf{b}_s\right)\cap P\left(y_{it}=1\right)}{P\left(y_{it}=1\right)} \end{aligned}$$

where we use $\widehat{p}_{it}$ as our estimate of the unconditional probability $P\left(y_{it}=1\right)$.

We could now decompose our binary choice model as follows:

$$\overline{p}_t-\overline{p}_s=\left(\overline{p}_t-\overline{p^a_{it}}\right)+\left(\overline{p^a_{it}}-\overline{p}_s\right) \tag{17}$$

The first term captures the effect of changes in the coefficients, keeping the characteristics as well as the unobserved errors constant, while the second effect captures both changes in the explanatory variables and in the error terms. Unlike with the linear model (where the decomposition in equation 7 is numerically equal to that in equation 3) this will not be precisely equal to that in equation 11. This follows since, in general $\overline{F\left(\mathbf{x}_{it}\mathbf{b}_s+\widehat{u}_{it}\right)}\neq\overline{F\left(\mathbf{x}_{it}\mathbf{b}_s\right)}$ even if $\overline{\widehat{u}_{it}}=0$. In practice Monte Carlo simulations suggest that the two means are very close to each other. This is not altogether surprising given that the cdf of the logistic and normal distributions is approximated reasonably well by a linear function in the range $0.3\leq p\leq 0.7$.

Given that the decompositions may often be similar, what do we gain from this more complicated procedure?

- In the first place one may **conceptually** prefer the decomposition given in equation 17. This will depend somewhat on how one views the process that generates the observations y_{it}. If one thinks that it is essentially a Bernoulli one (with parameter p as given in equation 9) then one might prefer the decomposition 11. In this view the outcome really is random and individuals that have the same $\mathbf{x}_{it}$ value are essentially interchangeable. If we were to re-run the social process, the same individual might flip the coin differently and end up with a different outcome. The latent variable model given in equations 13a and 13b by contrast assumes that the process is not really random. There are unobservable determinants which would induce the individual to act in very similar ways if the process were repeated. In particular if $\mathbf{b}_s=\mathbf{b}_t$ the outcome would be identical. The decomposition given in equation 17 attempts to freeze those unobservables when we change the coefficients (or indeed the characteristics).

- Secondly there is no guarantee that the means of $\widehat{p}_{it}^{o}$ and p_{it}^{a} would be similar over subpopulations. This could be due to two effects. Some subpopulations may have mean probability values well away from those characterising the sample as a whole, so that the non-linearity of the function F in that region may accentuate the difference between $\overline{F(\mathbf{x}_{it}\mathbf{b}_s + \widehat{u}_{it})}$ and $\overline{F(\mathbf{x}_{it}\mathbf{b}_s)}$. Furthermore there may be unobserved heterogeneity, so that some subpopulation errors may deviate systematically from the posited model. In the empirical analysis below we will show that there can be quite large divergences between the subpopulation means.

- Indeed, if the model is misspecified in the sense that the error structure is not precisely normal or logistic, we know that the Maximum Likelihood Estimator will still provide the "best" estimates of that misspecified model. Conditioning on the actual y_{it} values seems to make the counterfactual estimates yet more robust. Since the residuals will absorb any misspecification, taking these into account in constructing the counterfactual seems to protect somewhat against the impact of misspecification – particularly if the analysis is extended to subsamples where we suspect heterogeneity may become an issue. Indeed in the exercise below we will attempt to look at the hypothetical values in quite small subsamples (age cohorts).

- The main attraction of the Lemieux procedure for continuous dependent variables is that the y_{it}^{a} values can be used to decompose the variance or other higher order functions of the entire distribution. This defence is less compelling in this case, since the higher order moments of a Bernoulli variable are simply functions of the parameter p. A decomposition of the mean is therefore all that seems to be required. Nevertheless the procedure outlined above can be adapted for multinomial models (with some difficulties – see below) and extensions of these that model heteroscedasticity in the unobserved error term. In these cases being able to model separately the impact of changes in the coefficients, changes in the characteristics and changes in the distribution of the error terms might potentially be useful.

In order to round off this discussion we present an alternative approach to decomposing a binary choice model by analogy with that presented in table 1. This approach is given in table 3.

$\overline{p}_t = \sum_i \omega_{it} y_{it}$	The sample proportion in period t
$\overline{p}_t^{a} = \sum_i \omega_{it} p_{it}^{a}$	The hypothetical proportion that we would observe in period t if we were to fix the characteristics (including unobserved ones) but allow the coefficients to change
$\overline{p}_t^{c} = \sum_i \omega_{it}^{a} y_{it}$	The hypothetical proportion that we would observe in period t if we were to fix the coefficients but allow the covariates to change
$\overline{p}_t^{ac} = \sum_i \omega_{it}^{a} p_{it}^{a}$	The hypothetical proportion that we would observe in period t if we were to allow the covariates and the coefficients to change

Table 3: Extended decomposition of a binary variable

4 Dissecting the post-apartheid labour market changes

4.1 Comparing different decompositions

Before we apply the decomposition in a more comprehensive way, we will examine how different versions perform in tracking the change between just two years, 1995 and 2004. For this purpose we will use a specification in which the explanatory variables are a quadratic in age; a linear spline in education, with knots at 3, 7 and 12 years of schooling, corresponding to completion of junior primary, senior primary and high school; dummies for province; a dummy for urban/rural and the household size. This specification is similar to unemployment or employment probits estimated in the literature (Kingdon and Knight 2004a). We did not include additional household controls, because of their likely endogeneity. The household size variable was included because of the discontinuous change in this variable across the data sets, noted earlier.

For the dependent variable we looked at three mutually exclusive states: employed, unemployed (on the strict definition) and not economically active. This requires some comment, since a number of researchers have argued that the "broad" definition is more appropriate in South Africa (Kingdon and Knight 2006). We preferred the strict definition for two reasons. Firstly, it is now the official definition of unemployment used in South Africa. Secondly, there is considerable evidence that the boundary between the non-searching unemployed and the not economically active is just as porous as the boundary between the searching unemployed and the "discouraged" (Dinkelman and Pirouz 2002). Indeed leaving out the non-participants from an analysis of the labour market seems wrong in theory and in practice (Wittenberg 2002). Just as some of the "discouraged" workers are likely to start searching if the probability of success increases, many of the non-participants will become active. Our sample of analysis is African males aged 16 to 65 and we calculate the proportions over this population. We use 2004 as our base and investigate what this sample would have looked like if the conditions of 1995 had obtained or if the sampled individuals had had the characteristics of their 1995 counterparts.

As Table 4 shows, the proportions as measured in the Household Surveys changed markedly over this period. The proportion employed decreased by around four percentage points, while the proportion unemployed almost doubled. The implied unemployment **rate** was 29.4 % in 2004 compared with 16.7% in 1995. In 2004 there was also a higher participation rate. In order to decompose these changes we use four approaches. Firstly, we use a standard approach to construct the counterfactual proportion $\overline{\widehat{p}}_{it}^{o}$ by means of a logit model. Unlike the standard Even-MacPherson or Nielsen decomposition, we do not use an expression like $\left(\overline{\widehat{p}}_{t}^{o} - \overline{\widehat{p}}_{s}\right)$ to investigate the impact of changes in the characteristics while keeping the coefficients constant (see equation 11). Instead we do so by means of a Lemieux-style reweighting, i.e. we take the actual y_{it} values and reweight them to make the 2004 distribution look like 1995. We "reconstruct" what the actual distribution in 1995 ought to have been like (given our model) by reweighting the counterfactual terms $\widehat{p}_{it}^{o}$ to the 1995 distribution. In short we do an analogous decomposition to that proposed in table 3, except using $\widehat{p}_{it}^{o}$ instead of p_{it}^{a}.

The second approach is that discussed in section 3.2 and summarised in table 3, i.e. we use a logit model to construct the counterfactual proportions, but correct these for the unobservables. The third approach is to use a multinomial logit model to construct the counterfactual proportion. To our knowledge this approach has not been used in any applied work, but the idea is straightforward and would be a simple extension of the Even-MacPherson type of decomposition. Again we complete the decomposition by means of the reweighting procedure. The fourth and final approach is the decomposition of a multinomial logit model discussed in appendix A.

Before discussing the results of the comparison, it is useful to reflect on how coherent the underlying modelling strategy is. Unemployment or employment probits have been estimated by many authors before. Such models assume that the outcome (i.e. the individual is unemployed or employed) is due to some choice mechanism. Whether the estimated coefficients reflect worker or employer preferences cannot be decided on the data (Kingdon and Knight 2004a, p.207). Nevertheless the issue is likely to be somewhat more complicated, since the individual's choices can be seen as the choice to search or not, and then the choice to accept a job offer, or not. Conditional on the individual being economically active and being in the application pool for a particular job, the employer can choose whether or not to make a job offer to the individual concerned. To the extent to which individuals understand the preference functions of potential employers, this will feed back into both their search and job acceptance decisions. The economic costs and benefits (in particular the wage offered and the costs of search) are likely to feature prominently in these. This suggests that simple employment, unemployment or non-participation models are likely to distort the underlying social process. Even if we view our approach as estimating a "reduced form" model, in which the explanatory variables proxy for the benefits or costs, flattening the decision process into a simple dichotomous one is likely to misrepresent the social mechanisms. Furthermore estimating separate employment, unemployment and non-participation models seems dubious, given that they are part of the same process. Indeed estimating these separately does not impose the constraint that the imputed values should add up to 1.

Estimating the proportions by means of a multinomial logit model improves in this regard, since the proportions will add up correctly. The multinomial logit model is unlikely to be an appropriate model either, since the underlying assumption of the irrelevance of independent alternatives will not be met. The "choice" between unemployment and employment is not independent of the availability of non-participation as an option. A better choice would be the nested logit model or one of the other extensions of the multinomial logit model (Bhat 1995, Greene, Hensher and Rose 2006, Hensher and Greene 2002). Unfortunately it is beyond the scope of this paper to extend the decomposition methodology to these models. We defer this to future work. Within the constraints of tractable models, the multinomial logit is likely to be an improvement on the simple logit specification.

Turning to the results in table 4 we present in the first column (labelled $\boldsymbol{\beta}$) the counterfactual proportion if the 2004 sample was given the coefficients applying in 1995. In the first row, for instance, we see that 49.9% of African males would have been employed if they had answered the 1995 questionnaire under 1995 economic conditions. The second column (labelled $\mathbf{x}$) gives the proportion that would have been employed if we reweight the 2004 sample to make it look like the 1995 sample. We see that 36.6% of African males would have been employed if the 2004 sample had had the same characteristics as the 1995 one. The third column $(\mathbf{x}, \boldsymbol{\beta})$ gives the proportion if we change both the characteristics and the propensity to be employed to their 1995 levels. The first row shows that using the uncorrected logit model we would expect the 1995 proportion to be 0.435. The actual proportion was 0.464, so the change that was due to other factors (given in the "residual" column) was 0.0284.

Impact of changing:	$\boldsymbol{\beta}$	$\mathbf{x}$	$\mathbf{x}, \boldsymbol{\beta}$	**residual**
Employed 0.427 in 2004; 0.464 in 1995				
$\overline{\widetilde{p}^{o}_{it}}$, logit model	0.499	0.366	0.435	0.0284
$\overline{p^{a}_{it}}$, corrected logit model	0.501	0.366	0.443	0.0205
$\overline{\widetilde{p}^{o}_{ijt}}$, MNL model	0.499	0.366	0.435	0.0284
$\overline{p^{a}_{ijt}}$, corrected MNL model	0.503	0.366	0.444	0.0196
Unemployed 0.178 in 2004; 0.093 in 1995				
$\overline{\widetilde{p}^{o}_{it}}$, logit model	0.085	0.174	0.091	0.0017
$\overline{p^{a}_{it}}$, corrected logit model	0.084	0.174	0.086	0.0064
$\overline{\widetilde{p}^{o}_{ijt}}$, MNL model	0.083	0.174	0.088	0.0047
$\overline{p^{a}_{ijt}}$, corrected MNL model	0.083	0.174	0.085	0.0075
Non-participation 0.395 in 2004; 0.444 in 1995				
$\overline{\widetilde{p}^{o}_{it}}$, logit model	0.419	0.460	0.479	-0.0354
$\overline{p^{a}_{it}}$, corrected logit model	0.415	0.460	0.475	-0.0308
$\overline{\widetilde{p}^{o}_{ijt}}$, MNL model	0.418	0.460	0.477	-0.0331
$\overline{p^{a}_{ijt}}$, corrected MNL model	0.414	0.460	0.471	-0.0271

Table 4: Hypothetical proportions if 2004 data set had 1995 coefficients, characteristics or both; African Males aged 16 to 65

Comparing all the results in the table, we see that the corrected multinomial model has the smallest residual when modelling the proportion employed and the proportion not participating. It does relatively worst in reproducing the proportion unemployed. The reason for this is due to the adding up constraint. The modelled $(\mathbf{x}, \boldsymbol{\beta})$ proportions for the two logit models both exceed 1 in total. Being unconstrained by the requirement to produce coherent estimates across the three proportions, the simple logit model does reasonably well in modelling the proportion unemployed. If we calculate the mean square error across all three categories, the corrected multinomial model does better than any of the other approaches with a root mean square error of 0.0198. The corrected logit model performs second best (RMSE is 0.0217) with the simple multinomial model third (0.0253) and the simple logit model in last position (0.0262). The moral of this story is that correcting the estimates for the unobservables is perhaps even more important than estimating with more elaborate models. Nevertheless the multinomial logit in this case also seems to be an improvement on the simple logit.

4.2 Analysing changes by age group

The case in favour of "correcting" the predicted values for unobservables is strengthened when we analyse the performance of the imputations on subpopulations. Since we are creating weights and hypothetical values at the individual level it is very easy to calculate the proportions not only in aggregate but over particular subgroups. In this case we are interested in looking at the proportion employed by age group. We show the results graphically in Figure 1. Several trends stand out. Firstly the uncorrected models give very similar results as do the two models correcting for the unobservables. Secondly the corrections have the effect of pulling the estimates upwards among twenty-five to thirty year olds and among fifty-five to sixty-five year olds. In both cases this has the effect of pulling the estimates closer to the true values. Thirdly some significant gaps between the imputed and the actual values remain, particularly around age thirty.

For the remainder of this paper we intend to work with both aggregate decompositions and decompositions by age. In order to ensure that the age profiles that we report are not biased by our choice of a quadratic in age, we fit models in which we use separate dummies for every age and dummies for every educational level obtained. The additional controls are as before, i.e. we use provincial dummies, an urban-rural dummy and household size.

4.3 The aggregate changes

In Figure 2 we show the hypothetical distributions (changing $\boldsymbol{\beta}$) while keeping the characteristics fixed at the March 2004 level. Because of the number of lines in this picture, we smoothed them to make the graph clearer[2]. Three data sets stand out:

- the 1993 PSLSD data set seems to pick up more employment at young ages than any of the other surveys;
- the 1995 OHS shows higher levels of employment after age 25 than any of the other data sets, except for the PSLSD. In fact at the peak employment level around age 40, this data set seems to pick up around 8 percentage points more employment than the others.
- the 2000 September LFS seems to pick up more employment among workers over 55 years than the other surveys do

Except for these three data sets the other ones are roughly in a band, but with appreciable movement between years. It is noticeable that the 2004 figures tend to lie right at the bottom of this band, which suggests that the propensity to be measured to be employed is lower in the 2004 data sets than in the others. Even discounting the evident outliers in 1993 and 1995 the picture suggests that the propensity to be measured to be employed has decreased over these data sets.

Figure 3 provides the corresponding picture when we reweight the 2004 data set to take on the characteristics of the earlier data sets. What is noticeable in this instance is the smooth upward progression in the employment profiles. Some of the sample populations in the earliest data sets (in particular 1993 and 1995) have characteristics which would have dramatically lowered the employment rate if these individuals had faced the conditions pertaining in the 2004 sample. The story here seems equally clear: the African male population has become more employable over time.

The aggregate trends for all three states is given in Figure 4. Even accounting for the fact that the 1993, 1995 and 2000 propensity to be employed (dotted line) seems too high, there seems to be a definite downward trend in that propensity over the period. The dashed line, by contrast, confirms the fact that the characteristics have changed in such a way that the population has become more employable. The net effect (correcting 1993 and 1995 down) seems to be to leave the employment level more or less unchanged. Given population growth this must imply net job creation, but not at a rate faster than the population growth rate.

[2]The unsmoothed graph is available from the author on request.

The non-participation graph suggests that the propensity to be a non-participant has decreased over time while the characteristics have changed in such a way to accentuate this trend. Consequently there has been a clear increase in participation. This must imply an increase in unemployment, which is confirmed in the bottom left panel of Figure 4. In the bottom right panel we calculate the implied unemployment rate given the proportion employed and the proportion unemployed. The coefficients (economic conditions as well as the measurement process) have changed in such a way over the years that the unemployment rate has tended to rise. The characteristics of the population (increased employability) have changed in such a way that we might have expected the unemployment rate to fall slowly. The "coefficients" effect, however, seems to have outweighed the effect of better characteristics. Consequently the measured unemployment rate has increased since 1993.

A final piece of evidence is contained in figure 5 which graphs the root mean square error of the decomposition for each of the data sets. It is evident that the decomposition was relatively least successful in two of the "odd year" October Household Surveys. Many comparative analyses have been based on the 1995, 1997 and 1999 OHSs, because they have been thought to be more comparable, given the fact that they have similar sample sizes. This figure suggests that the neglect of the even years (1994, 1996 and 1998) may be unwarranted.

5 Conclusion

Several points emerge from the empirical analysis. Firstly, there are year-on-year shifts in the propensities which seem too large to be real. We suspect that the 1993, 1995 and 2000 data sets found too much employment and in the case of the last data set too little non-participation. The problematic nature of the 1995 data set is particularly noteworthy, given how central this data set has been to previous discussions of post-apartheid trends (Branson and Wittenberg 2006). Secondly, despite these concerns the decompositions paint a plausible picture overall:

- Changes in the characteristics of the population since 1993 (such as in education and location) have made African males more employable, more likely to participate and on balance should have brought the unemployment rate down
- Changes in economic conditions and in the measurement process have reduced the propensity to be employed, increased the propensity to be economically active and (on both counts) increased the propensity to be unemployed.
- The actual changes are the outcome of these some times opposing and some times complementary tendencies. In the case of employment, the total change is the net effect of two offsetting trends. Discounting data errors in the early period, the underlying change is likely to have been either flat or perhaps a small increase over time. Given population growth this would translate into some real employment gains, although not on the scale required to address poverty. In the case of participation, however, both trends work in the same direction, leading to a marked increase in the participation rate. As a result, the unemployment rate has undoubtedly increased.

What might these trends say about the performance of the post-apartheid labour market? The fact that employment prospects have worsened may be linked to the increase in participation.

As more people join the labour force competition for jobs will become heightened, unless demand increases even more. Our evidence suggests that there has at best been moderate employment growth. Why then the marked increase in participation? What would induce people to join the labour force in such numbers if the probability of finding a job is constantly diminishing? Several explanations come to mind:

- Perhaps the simplest is the one advanced by Branson and Wittenberg (2006). They suggest that changes in the schooling system have led to much faster exit rates from schools among black South Africans. These post-apartheid cohorts seem to achieve the same (or even better) educational outcomes as earlier ones, but leave the schools faster.
- It is possible that certain categories of individuals have become active in the labour market precisely because of the rising unemployment rate. This "added worker effect" (Lundberg 1985) was raised as a possible explanation for the increased participation by South African women (Casale and Posel 2002). This point would be harder to reconcile with the fact that we are also showing higher participation rates by African men.
- The lifting of all restrictions to access to jobs might have released a pent-up demand for participation that was only revealed with the demise of apartheid.

One might be tempted to assume that the value of "outside options" might have decreased over this period. This is not consistent with the fact that social welfare payments increased substantially, which should have increased the value of non-participation. It might, of course, also have increased the value of search.

In this paper we will not be able to settle these issues. Indeed our main contribution is that our decomposition stays clear of some of the concerns about the comparability of the data sets that has stymied some of the debates. Although the analysis is not immune to these problems, it does offer a more systematic way of disentangling changes in the characteristics of the population (real or due to changes in sampling design) from changes in the propensity to be measured in these states.

Indeed another important contribution of this paper has been to develop a Lemieux-style decomposition for discrete data. We initially showed for a binary choice model how we could construct individual level hypothetical data imposing the response rates from another survey, while keeping the effect of unobservables at their original level. In the appendix we extended this approach to a multinomial model with three categories. The approach could be adapted for higher order models, but would become rapidly intractable.

In the empirical part of the paper we showed that correcting for the unobservables does make a difference. It improves the accuracy of the decomposition (as measured by the root mean square error) and improves the fit of the relationship on various subsamples. On balance, the decomposition technique seems to work well. A particular attraction is that graphs like figures 2 and 3 show what happens in particular subpopulations. The standard numerical decomposition techniques do not lend themselves to this type of analysis as easily. Furthermore the standard techniques amalgamate changes in characteristics and changes in the residuals. The reweighting approach of diNardo et al. (1996) allows us to pick these apart. Nevertheless our results also have their limitation. A key problem is that the decomposition is only as good as the underlying model.

To the extent to which the multinomial model distorts the social process, our decomposition will also be incorrect.

Nevertheless dealing with the unobservables in the decomposition seems a significant step forward. The hypothetical experiment of subjecting individual A observed in year Y to the conditions of another year, viz. X , is more convincing if we can in some way acknowledge all the myriad ways in which traits of A, other than the ones that we had the fortune to measure, might have mattered for the labour market outcome.

A Decomposing a proportion based on a random utility model

In order to extend the approach outlined above to a multinomial logit model, we consider a random utility model. Assume that individual i has choices $0, \ldots, J$, i.e. $y_i \in \{0, \ldots, J\}$ where the utility of choice j is given by

$$U_{ij} = \mathbf{x}_i \boldsymbol{\beta}_j + \varepsilon_{ij} \tag{18}$$

We assume that j is chosen if $U_{ij} > U_{ik}$ for **all** $k \neq j$, i.e. if

$$\varepsilon_{ik} < \mathbf{x}_i \left(\boldsymbol{\beta}_j - \boldsymbol{\beta}_k \right) + \varepsilon_{ij} \text{ for all } k \neq j \tag{19}$$

If the ε_{ij} terms are independently distributed, then we can write the unconditional probability that option j is chosen as

$$\Pr(y_i = j) = \int_{\varepsilon_{ij}=-\infty}^{\varepsilon_{ij}=\infty} \prod_{k \neq j} F_k \left(\mathbf{x}_i \left(\boldsymbol{\beta}_j - \boldsymbol{\beta}_k \right) + \varepsilon_{ij} \right) f_j (\varepsilon_{ij}) \, d\varepsilon_{ij}$$

where F_j and f_j are the cdf and pdf of the ε_j terms. As McFadden showed, if the ε_{ij} terms are iid with extreme value distribution of type I, then the resulting unconditional probabilities take the form

$$\Pr(y_i = j) = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{k=0}^{J} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} \tag{20}$$

This is the multinomial logit model and J of the parameter vectors can be estimated by maximum likelihood. The standard procedure is to set $\boldsymbol{\beta}_0 = \mathbf{0}$. In the case where $J = 1$ this reduces to the standard logit model.

We can now consider the situation where we observe the choices in two time periods t and s, as before. As before we can estimate these coefficient vectors in each time period and contemplate what the effect would be if an individual in time t where to face the social process described by the parameter vectors $\mathbf{b}_{1s}, \ldots, \mathbf{b}_{Js}$. In the case of the multinomial logit model we get

$$\widehat{p}_{ijt}^{o} = \frac{e^{\mathbf{x}_{it} \mathbf{b}_{js}}}{1 + \sum_{k=1}^{J} e^{\mathbf{x}_{it} \mathbf{b}_{ks}}} \tag{21}$$

We can now consider how this counterfactual probability might shift if we fix the residuals at their level t values. In particular, assume that we know that in period t option l was chosen. In terms of our model we know that

$$\widehat{\varepsilon}_{ik} < \mathbf{x}_{it}\left(\mathbf{b}_{lt} - \mathbf{b}_{kt}\right) + \widehat{\varepsilon}_{il} \text{ for all } k \neq l$$

Note that there are J inequalities but $J+1$ unobserved residuals. Consequently one of the residuals will not be restricted by these inequalities. Observe furthermore that the behaviour of this model depends only on the bivariate comparisons of the type $\mathbf{x}_{it}\left(\mathbf{b}_{lt} - \mathbf{b}_{kt}\right)$ or $\mathbf{x}_{it}\left(\mathbf{b}_{js} - \mathbf{b}_{ks}\right)$. These terms can be thought of as the deterministic part of the index for the comparison of the option l against k (in period t) or j against k (in period t using the coefficients from period s). Let $\upsilon_{lkt} = \mathbf{x}_{it}\left(\mathbf{b}_{lt} - \mathbf{b}_{kt}\right)$ and $\upsilon_{jks} = \mathbf{x}_{it}\left(\mathbf{b}_{js} - \mathbf{b}_{ks}\right)$ and similarly for the other possibilities. We have omitted reference to observation i, to economise on notation. Observe also that by definition $\upsilon_{jkt} = -\upsilon_{kjt}$.

We want to calculate

$$p^a_{ijt} = \Pr\left(\widehat{\varepsilon}_{i0} < \upsilon_{j0s} + \widehat{\varepsilon}_{ij} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{jJs} + \widehat{\varepsilon}_{ij} \left|\widehat{\varepsilon}_{i0} < \upsilon_{l0t} + \widehat{\varepsilon}_{il} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{lJt} + \widehat{\varepsilon}_{il}\right.\right)$$

This conditional probability can be written as the joint probability divided by the marginal probability. But this marginal probability is

$$\Pr\left(\widehat{\varepsilon}_{i0} < \upsilon_{l0t} + \widehat{\varepsilon}_{il} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{lJt} + \widehat{\varepsilon}_{il}\right) = p_{ilt}$$

which is the probability of outcome l estimated (for instance by the multinomial model) for period t. Since we have a convenient expression for the denominator, we need to consider only the joint probability

$$\Pr\left(\widehat{\varepsilon}_{i0} < \upsilon_{j0s} + \widehat{\varepsilon}_{ij} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{jJs} + \widehat{\varepsilon}_{ij} \text{ and } \widehat{\varepsilon}_{i0} < \upsilon_{l0t} + \widehat{\varepsilon}_{il} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{lJt} + \widehat{\varepsilon}_{il}\right) \tag{22}$$

We now use the fact that we can treat one of the residuals as unrestricted. We will condition on $\widehat{\varepsilon}_{ij}$, i.e. we will be able to write p^a_{ijt} as

$$p^a_{ijt} = \frac{1}{p_{ilt}} \int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty} \Pr\left(\begin{array}{c} \widehat{\varepsilon}_{i0} < \upsilon_{j0s} + \widehat{\varepsilon}_{ij} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{jJs} + \widehat{\varepsilon}_{ij} \text{ and } \\ \widehat{\varepsilon}_{i0} < \upsilon_{l0t} + \widehat{\varepsilon}_{il} \text{ and } \cdots \text{ and } \widehat{\varepsilon}_{iJ} < \upsilon_{lJt} + \widehat{\varepsilon}_{il} \end{array} \left|\widehat{\varepsilon}_{ij}\right.\right) f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \tag{23}$$

where f_j is the pdf of $\widehat{\varepsilon}_{ij}$. There are two cases to consider:

A.1 Case 1: Option j was chosen in period t

In this case the condition $\left(\widehat{\varepsilon}_{ik} < \upsilon_{jks} + \widehat{\varepsilon}_{ij} \text{ and } \widehat{\varepsilon}_{ik} < \upsilon_{jkt} + \widehat{\varepsilon}_{ij}\right)$ can be simplified to $\widehat{\varepsilon}_{ik} < \min\left\{\upsilon_{jks}, \upsilon_{jkt}\right\} + \widehat{\varepsilon}_{ij}$. Assuming that the $\widehat{\varepsilon}_{ik}$ terms are independently distributed, we can then write

$$p^a_{ijt} = \frac{1}{p_{ijt}} \int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty} \left\{\prod_{k \neq j} F_k\left(\min\left\{\upsilon_{jks}, \upsilon_{jkt}\right\} + \widehat{\varepsilon}_{ij}\right)\right\} f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \tag{24}$$

Two polar cases are noteworthy. If $v_{jks} \geq v_{jkt}$ for every $k \neq j$, then the probability simplifies to p_{ijt} and hence $p^a_{ijt} = 1$. If the inequalities are reversed, then it simplifies to $\widehat{p}^o_{ijt}$, i.e. $p^a_{ijt} = \widehat{p}^o_{ijt}/p_{ijt}$. In the case where $J = 1$ these are the only two possibilities, hence we get the same result that we got earlier.

Assume now that there are exactly three options. Let us label these j, k and l. Furthermore let $v_k = \min\{v_{jks}, v_{jkt}\}$ and assume that the error terms have extreme value distribution of type I. Then equation 24 becomes

$$\begin{aligned} p^a_{ijt} &= \frac{1}{p_{ijt}} \int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty} \left\{\exp\left(-e^{-v_k-\widehat{\varepsilon}_{ij}}\right) \exp\left(-e^{-v_l-\widehat{\varepsilon}_{ij}}\right)\right\} e^{-\widehat{\varepsilon}_{ij}} \exp\left(-e^{-\widehat{\varepsilon}_{ij}}\right) d\widehat{\varepsilon}_{ij} \\ &= \frac{1}{p_{ijt}} \frac{1}{1+e^{-v_k}+e^{-v_l}} \end{aligned} \tag{25}$$

It is easily verified that this formula will generalise in the case of more than three outcomes. It is also straightforward to show that if **every** minimum belongs to the same period (i.e. s or t) then the formula simplifies to either p_{ijt} or $\widehat{p}^o_{ijt}$, as discussed above.

A.2 Case 2: Option $l \neq j$ was chosen in period t

In this case we require both $\widehat{\varepsilon}_{il} < v_{jls} + \widehat{\varepsilon}_{ij}$ and $\widehat{\varepsilon}_{ij} < v_{ljt} + \widehat{\varepsilon}_{il}$, i.e. we require

$$-v_{ljt} + \widehat{\varepsilon}_{ij} < \widehat{\varepsilon}_{il} < v_{jls} + \widehat{\varepsilon}_{ij}$$

If this condition cannot be met for any values of $\widehat{\varepsilon}_{ij}$ then the joint probability given in equation 22 must be zero. In order for the probability to be non-zero we require $v_{jlt} < v_{jls}$. This means that j must become relatively more attractive in period s than it was in period t if there is to be a non-zero probability of choosing it. This is obvious: if option l has become relatively more attractive in period s and it was already chosen in period t, then in terms of utility maximisation j will not be chosen in period s either.

Assume now that $v_{jlt} < v_{jls}$. For $k \neq l$ and $k \neq j$ we require $\widehat{\varepsilon}_{ik} < v_{jks} + \widehat{\varepsilon}_{ij}$ and $\widehat{\varepsilon}_{ik} < v_{lkt} + \widehat{\varepsilon}_{il}$, i.e. $\widehat{\varepsilon}_{ik} < \min\{v_{jks} + \widehat{\varepsilon}_{ij}, v_{lkt} + \widehat{\varepsilon}_{il}\}$. We assume again that the $\widehat{\varepsilon}_{ik}$ terms are independently distributed. To evaluate the probability in equation 23, we will condition now also on $\widehat{\varepsilon}_{il}$, i.e. we will have

$$p^a_{ijt} = \frac{1}{p_{ilt}} \int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty} \int_{\widehat{\varepsilon}_{il}=v_{jlt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=v_{jls}+\widehat{\varepsilon}_{ij}} \left\{\prod_{k\neq j,l} F_k\left(\min\{v_{jks} + \widehat{\varepsilon}_{ij}, v_{lkt} + \widehat{\varepsilon}_{il}\}\right)\right\} f_l\left(\widehat{\varepsilon}_{il}\right) f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{il} d\widehat{\varepsilon}_{ij} \tag{26}$$

The behaviour of the terms in the inner-most integral turn out to depend on the changes in the relative attractiveness of j versus k and l versus k between the two periods. We will again restrict our attention to the multinomial logit model with $J = 3$. In this case there are four possibilities depending on the magnitude of v_{jkt} relative to v_{jks} and v_{lkt} relative v_{lks}. These give four expressions for p^a_{ijt}, as shown in Table 5. The derivations are somewhat tedious, so are provided in a separate appendix below.

Generalising these results to higher dimensional problems is difficult, since the possibilities start multiplying rapidly. In particular for every additional variable k that fits into the bottom

	$\upsilon_{lkt} < \upsilon_{lks}$	$\upsilon_{lkt} > \upsilon_{lks}$
$\upsilon_{jkt} > \upsilon_{jks}$	$p^a_{ijt} = 0$	$\frac{1}{p_{ilt}}\left(\widehat{p}^o_{ijt} - \frac{1}{1+e^{\upsilon_{kjs}}+e^{\upsilon_{ljt}}}\right)$
$\upsilon_{jkt} < \upsilon_{jks}$	$(1-p_{ijt})^{-1}\left(\frac{e^{\upsilon_{jls}}}{1+e^{\upsilon_{jls}}+e^{\upsilon_{klt}}} - p_{ijt}\right)$	$\frac{\widehat{p}^o_{ijt}}{p_{ilt}} - \frac{1}{p_{ilt}}\frac{e^{\upsilon_{klt}}}{1+e^{\upsilon_{klt}}}\frac{e^{\upsilon_{jks}}}{1+e^{\upsilon_{jks}}+e^{\upsilon_{lkt}}} - \frac{1}{p_{ilt}}\frac{1}{1+e^{-\upsilon_{lkt}}}p_{ijt}$

Table 5: Values for p^a_{ijt} in the multinomial logit model with $J = 3$

right quadrant of this typology the domain of the innermost integral of equation 26 will have to be subdivided further.

The term $\frac{e^{\upsilon_{jks}}}{1+e^{\upsilon_{jks}}+e^{\upsilon_{lkt}}}$ that occurs in several of these expressions deserves further comment. We have

$$\frac{e^{\upsilon_{jks}}}{1+e^{\upsilon_{jks}}+e^{\upsilon_{lkt}}} = \frac{e^{\mathbf{x}_i\mathbf{b}_{js}}}{e^{\mathbf{x}_i\mathbf{b}_{ks}} + e^{\mathbf{x}_i\mathbf{b}_{js}} + e^{\mathbf{x}_i\mathbf{b}_{lt}}e^{\mathbf{x}_i(\mathbf{b}_{ks}-\mathbf{b}_{kt})}}$$

Except for the term $e^{\mathbf{x}_i(\mathbf{b}_{ks}-\mathbf{b}_{kt})}$ in the denominator, this looks like a multinomial probability, where the coefficients on choices j and k have been changed to the values obtaining in year s, but the coefficient on choice l has been frozen at the year t level. Note that if k happens to be the baseline case then $\mathbf{b}_{ks} = \mathbf{b}_{kt} = 0$, i.e. the expression would correspond exactly to a probability with index values $\mathbf{x}_i\mathbf{b}_{js}$, $\mathbf{x}_i\mathbf{b}_{ks}$ and $\mathbf{x}_i\mathbf{b}_{lt}$. The term $\frac{e^{\upsilon_{jks}}}{1+e^{\upsilon_{jks}}+e^{\upsilon_{lkt}}}$ can therefore be thought of as the unconditional probability that y_{it} is equal to j using k as the base case and keeping the relative attractiveness of option l at its level in period t. The term $e^{\mathbf{x}_i(\mathbf{b}_{ks}-\mathbf{b}_{kt})}$ functions as a correction – when period s index values are compared to period t ones they are both scaled implicitly by the condition that the coefficient vector on the base case should be identically zero in both periods. This highlights the fact that if coefficients from different periods are "mixed and matched" the implied probabilities will not be invariant to the choice of the base case.

B Derivations

We present the derivations in the order given in Table 5 starting at the top left hand corner. We note that if $\upsilon_{jlt} > \upsilon_{jls}$ then $p^a_{ijt} = 0$, so we assume that $\upsilon_{jlt} < \upsilon_{jls}$. Furthermore we require

$$\upsilon_{jlt} + \widehat{\varepsilon}_{ij} < \widehat{\varepsilon}_{il} < \upsilon_{jls} + \widehat{\varepsilon}_{ij}$$

- **Possibility 1:** $\upsilon_{jkt} > \upsilon_{jks}$ and $\upsilon_{lkt} < \upsilon_{lks}$

  These assumptions are $\mathbf{x}_i(\mathbf{b}_{jt} - \mathbf{b}_{kt}) > \mathbf{x}_i(\mathbf{b}_{js} - \mathbf{b}_{ks})$ **and** $\mathbf{x}_i(\mathbf{b}_{lt} - \mathbf{b}_{kt}) < \mathbf{x}_i(\mathbf{b}_{ls} - \mathbf{b}_{ks})$ These conditions imply that $\mathbf{x}_i(\mathbf{b}_{jt} - \mathbf{b}_{lt}) > \mathbf{x}_i(\mathbf{b}_{js} - \mathbf{b}_{ls})$, i.e. $\upsilon_{jlt} > \upsilon_{jls}$

- **Possibility 2:** $\upsilon_{jkt} > \upsilon_{jks}$ and $\upsilon_{lkt} > \upsilon_{lks}$

  We have $\upsilon_{lkt} + \widehat{\varepsilon}_{il} > \upsilon_{lkt} + \upsilon_{jlt} + \widehat{\varepsilon}_{ij} = \upsilon_{jkt} + \widehat{\varepsilon}_{ij}$. It follows that $\min\{\upsilon_{jks} + \widehat{\varepsilon}_{ij}, \upsilon_{lkt} + \widehat{\varepsilon}_{il}\} = \upsilon_{jks} + \widehat{\varepsilon}_{ij}$.

We now consider what this implies for p^a_{ijt} in the case where $J=3$. Equation 26 becomes

$$\begin{aligned}
p^a_{ijt} &= \frac{1}{p_{ilt}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\int_{\widehat{\varepsilon}_{il}=\upsilon_{jlt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=\upsilon_{jls}+\widehat{\varepsilon}_{ij}} F_k\left(\upsilon_{jks}+\widehat{\varepsilon}_{ij}\right) f_l\left(\widehat{\varepsilon}_{il}\right) f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{il} d\widehat{\varepsilon}_{ij} \\
&= \frac{1}{p_{ilt}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\left[F_k\left(\upsilon_{jks}+\widehat{\varepsilon}_{ij}\right)\left\{F_l\left(\upsilon_{jls}+\widehat{\varepsilon}_{ij}\right)-F_l\left(\upsilon_{jlt}+\widehat{\varepsilon}_{ij}\right)\right\}\right] f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \\
&= \frac{1}{p_{ilt}}\left\{\widehat{p}^o_{ijt}-\frac{1}{1+e^{-\upsilon_{jks}}+e^{-\upsilon_{jlt}}}\right\}
\end{aligned}$$

- **Possibility 3:** $\upsilon_{jkt}<\upsilon_{jks}$ and $\upsilon_{lkt}<\upsilon_{lks}$

  It follows that $\upsilon_{lkt}+\widehat{\varepsilon}_{il}<\upsilon_{lks}+\upsilon_{jls}+\widehat{\varepsilon}_{ij}=\upsilon_{jks}+\widehat{\varepsilon}_{ij}$. Consequently $\min\left\{\upsilon_{jks}+\widehat{\varepsilon}_{ij},\upsilon_{lkt}+\widehat{\varepsilon}_{il}\right\}=\upsilon_{lkt}+\widehat{\varepsilon}_{il}$.

  In this case

$$\begin{aligned}
p^a_{ijt} &= \frac{1}{p_{ilt}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\left[\int_{\widehat{\varepsilon}_{il}=\upsilon_{jlt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=\upsilon_{jls}+\widehat{\varepsilon}_{ij}} F_k\left(\upsilon_{jks}+\widehat{\varepsilon}_{il}\right) f_l\left(\widehat{\varepsilon}_{il}\right) d\widehat{\varepsilon}_{il}\right] f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \\
&= \frac{1}{p_{ilt}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\left[\frac{1}{e^{\upsilon_{lkt}}+1}\exp\left(\upsilon_{lkt}-e^{-\widehat{\varepsilon}_{il}}-e^{-\upsilon_{lkt}-\widehat{\varepsilon}_{il}}\right)\right]_{\widehat{\varepsilon}_{il}=\boldsymbol{\upsilon}_{jlt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=\boldsymbol{\upsilon}_{jls}+\widehat{\varepsilon}_{ij}} f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \\
&= \frac{1}{p_{ilt}}\frac{1}{1+e^{-\upsilon_{lkt}}}\left[\frac{e^{\upsilon_{jls}}}{1+e^{\upsilon_{jls}}+e^{-\upsilon_{lkt}}}-\frac{e^{\upsilon_{jlt}}}{1+e^{\upsilon_{jlt}}+e^{-\upsilon_{lkt}}}\right]
\end{aligned}$$

  Now note that $p_{ilt}=\frac{e^{\mathbf{x}_i\mathbf{b}_{lt}}}{e^{\mathbf{x}_i\mathbf{b}_{jt}}+e^{\mathbf{x}_i\mathbf{b}_{kt}}+e^{\mathbf{x}_i\mathbf{b}_{lt}}}$ while $\frac{1}{1+e^{-\upsilon_{lkt}}}=\frac{e^{\mathbf{x}_i\mathbf{b}_{lt}}}{e^{\mathbf{x}_i\mathbf{b}_{kt}}+e^{\mathbf{x}_i\mathbf{b}_{lt}}}$ and hence

$$\frac{1}{p_{ilt}}\frac{1}{1+e^{-\upsilon_{lkt}}}=\frac{e^{\mathbf{x}_i\mathbf{b}_{jt}}+e^{\mathbf{x}_i\mathbf{b}_{kt}}+e^{\mathbf{x}_i\mathbf{b}_{lt}}}{e^{\mathbf{x}_i\mathbf{b}_{kt}}+e^{\mathbf{x}_i\mathbf{b}_{lt}}}=\left(1-p_{ijt}\right)^{-1}$$

- **Possibility 4:** $\upsilon_{jkt}<\upsilon_{jks}$ and $\upsilon_{lkt}>\upsilon_{lks}$

  We observe that if $\widehat{\varepsilon}_{il}=\upsilon_{jlt}+\widehat{\varepsilon}_{ij}$ we must have $\upsilon_{lkt}+\widehat{\varepsilon}_{il}=\upsilon_{lkt}+\upsilon_{jlt}+\widehat{\varepsilon}_{ij}=\upsilon_{jkt}+\widehat{\varepsilon}_{ij}$ i.e. $\min\left\{\upsilon_{jks}+\widehat{\varepsilon}_{ij},\upsilon_{lkt}+\widehat{\varepsilon}_{il}\right\}=\upsilon_{lkt}+\widehat{\varepsilon}_{il}$ when $\upsilon_{jlt}+\widehat{\varepsilon}_{ij}<\widehat{\varepsilon}_{il}<\upsilon_{jks}-\upsilon_{lkt}+\widehat{\varepsilon}_{ij}$ and $\min\left\{\upsilon_{jks}+\widehat{\varepsilon}_{ij},\upsilon_{lkt}+\widehat{\varepsilon}_{il}\right\}=\upsilon_{jks}+\widehat{\varepsilon}_{ij}$ when $\upsilon_{jks}-\upsilon_{lkt}+\widehat{\varepsilon}_{ij}<\widehat{\varepsilon}_{il}<\upsilon_{jls}+\widehat{\varepsilon}_{ij}$. Hence

$$\begin{aligned}
p^a_{ijt} &= \frac{1}{p_{ilt}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\left\{\left[\begin{array}{c}\int_{\widehat{\varepsilon}_{il}=\upsilon_{jlt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=\upsilon_{jks}-\upsilon_{lkt}+\widehat{\varepsilon}_{ij}} F_k\left(\upsilon_{lkt}+\widehat{\varepsilon}_{il}\right) f_l\left(\widehat{\varepsilon}_{il}\right) d\widehat{\varepsilon}_{il} \\ +\int_{\widehat{\varepsilon}_{il}=\upsilon_{jks}-\upsilon_{lkt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=\upsilon_{jls}+\widehat{\varepsilon}_{ij}} F_k\left(\upsilon_{jks}+\widehat{\varepsilon}_{ij}\right) f_l\left(\widehat{\varepsilon}_{il}\right) d\widehat{\varepsilon}_{il}\end{array}\right]\right\} f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \\
&= \frac{1}{p_{ilt}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\left\{\begin{array}{c}\left[\frac{1}{e^{\upsilon_{lkt}+1}}\exp\left(\upsilon_{lkt}-e^{-\widehat{\varepsilon}_{il}}-e^{-\upsilon_{lkt}-\widehat{\varepsilon}_{il}}\right)\right]_{\widehat{\varepsilon}_{il}=\upsilon_{jlt}+\widehat{\varepsilon}_{ij}}^{\widehat{\varepsilon}_{il}=\upsilon_{jks}-\upsilon_{lkt}+\widehat{\varepsilon}_{ij}}+ \\ F_k\left(\upsilon_{jks}+\widehat{\varepsilon}_{ij}\right)\left\{F_l\left(\upsilon_{jls}+\widehat{\varepsilon}_{ij}\right)-F_l\left(\upsilon_{jks}-\upsilon_{lkt}+\widehat{\varepsilon}_{ij}\right)\right\}\end{array}\right\} f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \\
&= \frac{1}{p_{ilt}}\frac{1}{1+e^{-\upsilon_{lkt}}}\int_{\widehat{\varepsilon}_{ij}=-\infty}^{\widehat{\varepsilon}_{ij}=\infty}\left\{\begin{array}{c}\exp\left(-e^{-\boldsymbol{\upsilon}_{jks}+\upsilon_{lkt}-\widehat{\varepsilon}_{ij}}-e^{-\upsilon_{jks}-\widehat{\varepsilon}_{ij}}\right) \\ -\exp\left(-e^{-\boldsymbol{\upsilon}_{jlt}-\widehat{\varepsilon}_{ij}}-e^{-\upsilon_{lkt}-\boldsymbol{\upsilon}_{jlt}-\widehat{\varepsilon}_{ij}}\right)\end{array}\right\} f_j\left(\widehat{\varepsilon}_{ij}\right) d\widehat{\varepsilon}_{ij} \\
&\quad +\frac{1}{p_{ilt}}\left\{\widehat{p}^o_{ijt}-\frac{e^{\boldsymbol{\upsilon}_{jks}}}{1+e^{\boldsymbol{\upsilon}_{jks}}+e^{\boldsymbol{\upsilon}_{lkt}}}\right\} \\
&= \frac{\widehat{p}^o_{ijt}}{p_{ilt}}-\frac{1}{p_{ilt}}\frac{e^{\upsilon_{klt}}}{1+e^{\upsilon_{klt}}}\frac{e^{\boldsymbol{\upsilon}_{jks}}}{1+e^{\boldsymbol{\upsilon}_{jks}}+e^{\boldsymbol{\upsilon}_{lkt}}}-\frac{1}{p_{ilt}}\frac{1}{1+e^{-\upsilon_{lkt}}}\frac{e^{\boldsymbol{\upsilon}_{jlt}}}{1+e^{\boldsymbol{\upsilon}_{jlt}}+e^{\upsilon_{klt}}}
\end{aligned}$$

References

Banerjee, Abhijit, Sebastian Galiani, Jim Levinsohn, and Ingrid Woolard, "Why has unemployment risen in the new South Africa?," Working Paper 134, Center for International Development, Harvard University October 2006.

Bhat, Chandra R., "A heteroscedastic extreme value model of intercity travel mode choice," *Transportation Research Part B*, 1995, *29*, 471–483.

Bhorat, Haroon, "The October Household Survey, Unemployment and the Informal Sector: A Note," *South African Journal of Economics*, 1999, *67* (2), 320–326.

—, "The Post-Apartheid Challenge: Labour Demand Trends in the South African Labour Market, 1995-1999," Working Paper 03/82, Development Policy Research Unit, University of Cape Town 2003.

—, "Labour Market Challenges in the post-apartheid South Africa," *South African Journal of Economics*, 2004, *72* (5), 940–977.

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 1973, *8* (4), 436–455.

Borooah, Vani K. and Sriya Iyer, "The decomposition of inter-group differences in a logit model: Extending the Oaxaca-Blinder approach with an application to school enrolment in India," *Journal of Economic and Social Measurement*, 2005, *30*, 279–293).

Branson, Nicola and Martin Wittenberg, "The post-apartheid South African Labour Market: A Cohort Analysis of National Data Sets, 1994-2004," 2006. mimeo. School of Economics, University of Cape Town.

Casale, Daniela and Dorrit Posel, "The continued feminisation of the labour force in South Africa: An analysis of recent data and trends," *South African Journal of Economics*, 2002, *70* (1), 156–184.

—, Colette Muller, and Dorrit Posel, "'Two million net new jobs': A Reconsideration of the Rise in Employment in South Africa, 1995–2003," *South African Journal of Economics*, 2004, *72* (5), 978–1002.

diNardo, John, Nicole Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distributions of Wages, 1973–1992: A Semiparametric Approach," *Econometrica*, 1996, *64* (5), 1001–1044.

Dinkelman, Taryn and Farah Pirouz, "Individual, household and regional determinants of labour force attachment in South Africa: Evidence from the 1997 October Household Survey," *South African Journal of Economics*, 2002, *70* (5), 865–891.

Even, William E and David A Macpherson, "Plant Size and the Decline of Unionism," *Economics Letters*, 1990, *32*, 393–398.

Fairlie, Robert W., "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models," *Journal of Economic and Social Measurement*, 2005, *30*, 305–316.

Greene, William H., David A. Hensher, and John Rose, "Accounting for heterogeneity in the variance of unobserved effects in mixed logit models," *Transportation Research Part B*, 2006, *40*, 75–92.

Hensher, David A. and William H. Greene, "Specification and estimation of the nested logit model: alternative normalisations," *Transportation Research Part B*, 2002, *36*, 1–17.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skills," *Journal of Political Economy*, 1993, *101* (3), 410–442.

Kingdon, Geeta and John Knight, "Race and the Incidence of Unemployment in South Africa," *Review of Development Economics*, 2004, *8* (2), 198–222.

___ **and** ___ , "Unemployment in South Africa: The Nature of the Beast," *World Development*, 2004, *32* (3), 391–408.

___ **and** ___ , "The measurement of unemployment when unemployment is high," *Labour Economics*, 2006, *13*, 291–315.

Klasen, Stephan and Ingrid Woolard, "Levels, Trends and Consistency of Measured Employment and Unemployment in South Africa," *Development Southern Africa*, 1999, *16*, 3–35.

___ **and** ___ , "Surviving Unemployment without State Support: Unemployment and Household Formation in South Africa," 2000. mimeo. University of Munich and University of Port Elizabeth.

___ **and** ___ , "Unemployment and employment in South Africa, 1995-1997: Report to the Department of Finance," 2000. mimeo. University of Munich and University of Port Elizabeth.

Lemieux, Thomas, "Decomposing Changes in Wage Distributions: A Unified Approach," *Canadian Journal of Economics*, 2002, *35* (4), 646–688.

Lundberg, Shelly, "The Added Worker Effect," *Journal of Labor Economics*, 1985, *3* (1), 11–37.

Mbeki, Thabo, "Letter from the President: May Day Part III," *ANC Today*, May 2005, *5* (20). Accessed electronically at http://www.anc.org.za/ancdocs/anctoday/2005/at20.htm.

Melly, Blaise, "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 2005, *12*, 577–590.

Nielsen, Helena Skyt, "Discrimination and detailed decomposition in a logit model," *Economics Letters*, 1998, *61*, 115–120.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labour Markets," *International Economic Review*, 1973, *14* (3), 693–709.

PCAS, *Towards a Ten Year Review: Synthesis Report on Implementation of Government Programmes*, Pretoria: Policy Co-ordination and Advisory Services (PCAS), Presidency, South Africa, October 2003. Available at http://www.gcis.gov.za/docs/publications/10year.pdf.

Seekings, Jeremy and Nicoli Nattrass, *Class, Race and Inequality in South Africa*, University of KwaZulu-Natal Press, 2006.

Standing, Guy, John Sender, and John Weeks, *Restructuring the Labour Market: The South African Challenge*, International Labour Office, 1996.

White, Halbert, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 1982, *50* (1), 1–26.

Wittenberg, Martin, "Job Search in South Africa: A Nonparametric Analysis," *South African Journal of Economics*, 2002, *70* (8), 1163–1198.

___ **and Mark Collinson**, "Household transitions in rural South Africa during 1996-2003," *Scandinavian Journal of Public Health*, forthcoming.

Yun, Myeong-Su, "Decomposing differences in the first moment," *Economics Letters*, 2004, *82*, 275–280.

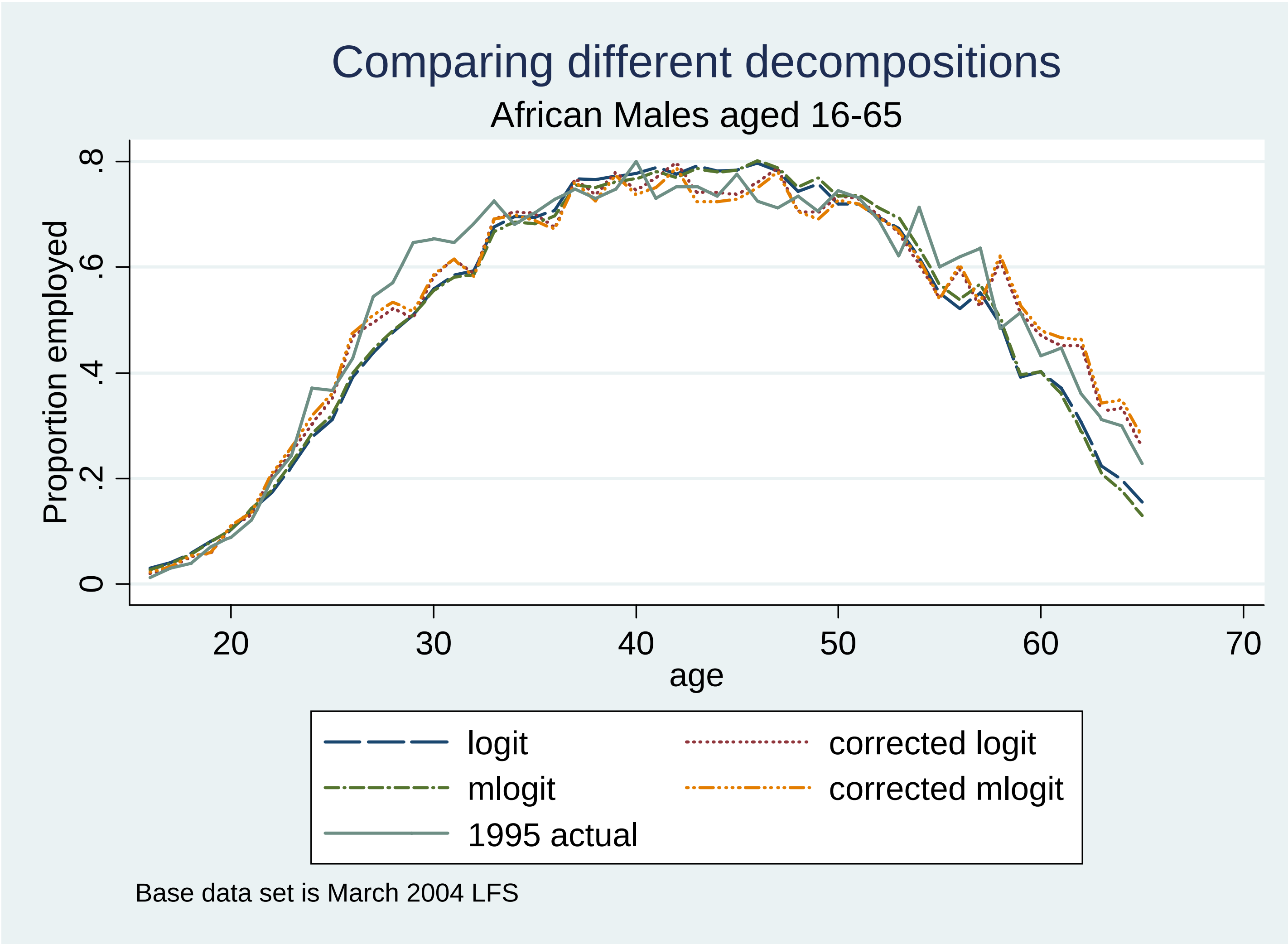

Figure 1: Predictions $(\mathbf{x}, \boldsymbol{\beta})$ changing both the characteristics and the coefficients of the 2004 sample to the 1995 levels

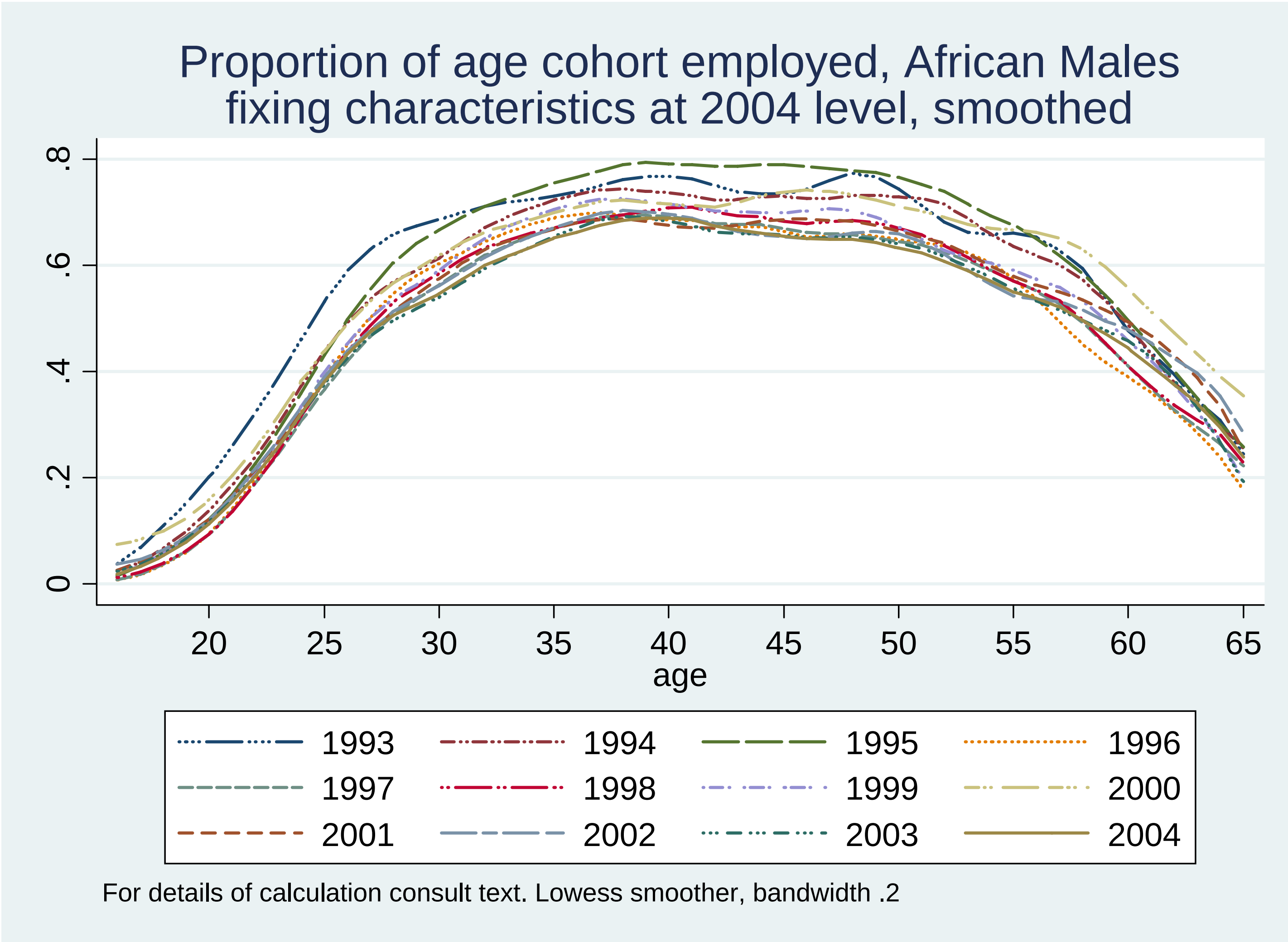

Figure 2: Comparing different data sets while keeping the characteristics constant and changing $\boldsymbol{\beta}$

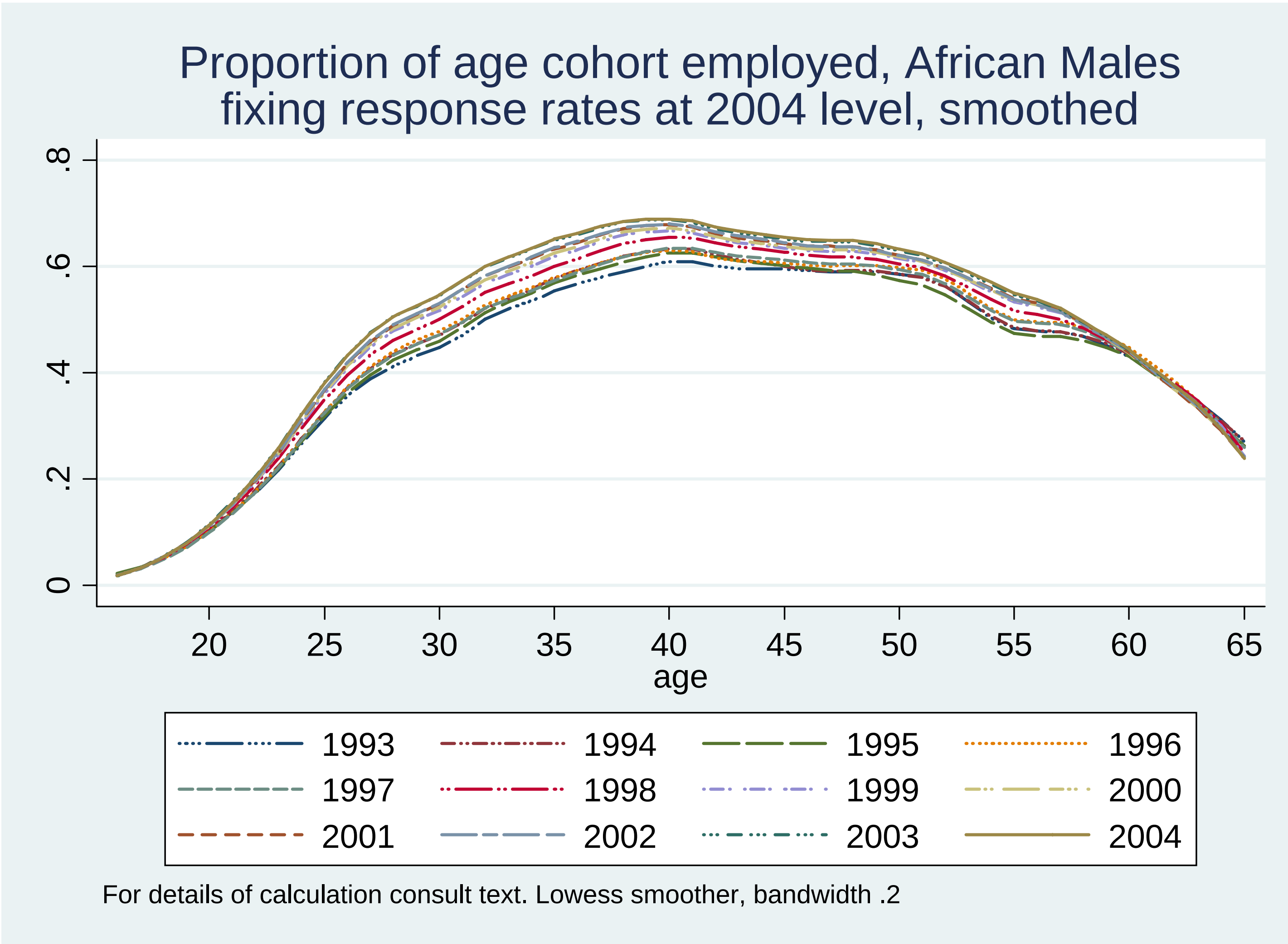

Figure 3: Comparing different data sets while keeping the coefficients constant and changing **x**

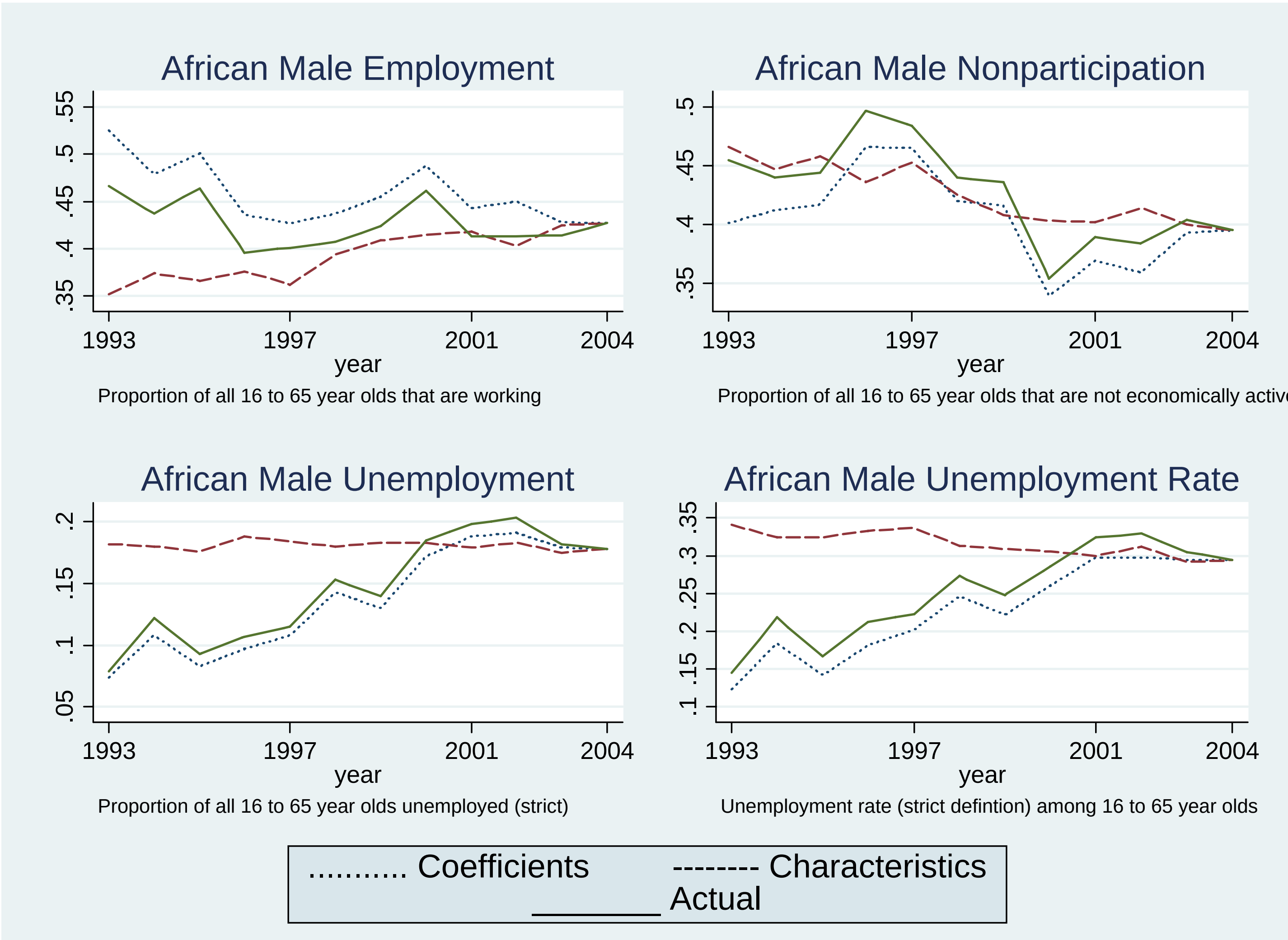

Figure 4: Aggregate labour market trends 1993-2004, African males aged 16 to 65

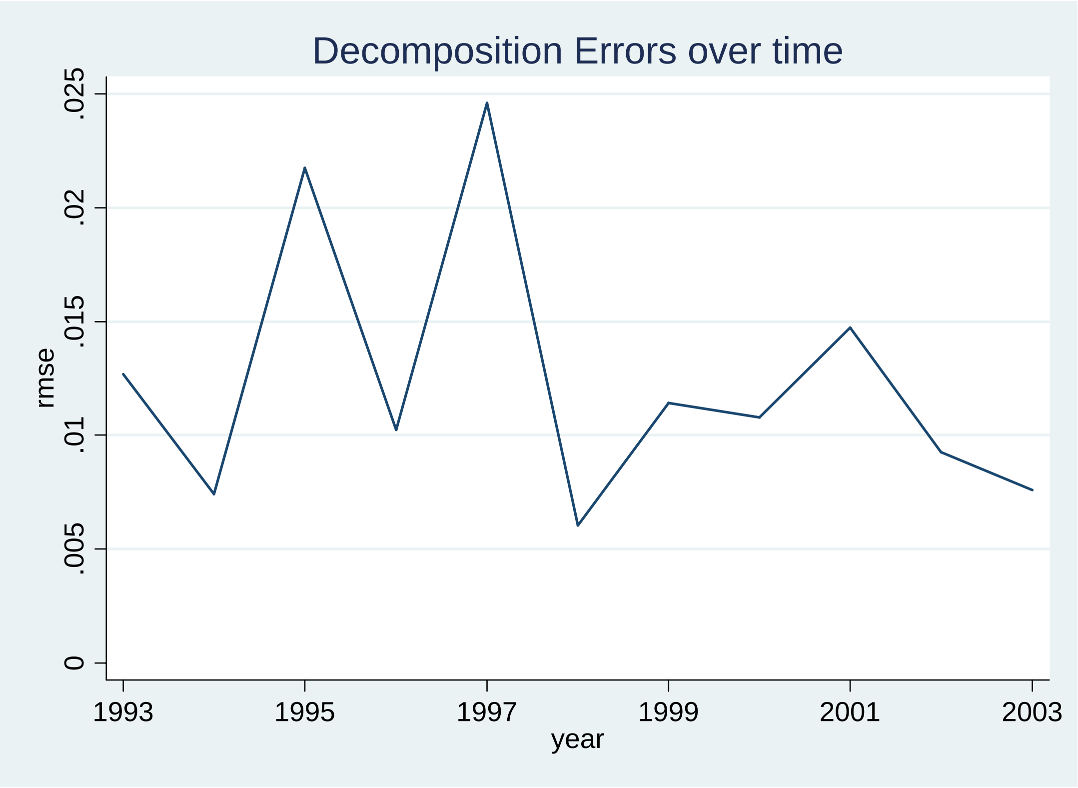

Figure 5: The average error in the decomposition across the three proportions was highest in 1995 and 1997