

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH

**RISING UNEMPLOYMENT IN SOUTH AFRICA: AN
INTERTEMPORAL ANALYSIS USING A BIRTH COHORT
PANEL**

by

Dieter von Fintel



**Assignment presented in partial fulfilment of the requirements for the degree of
Masters of Commerce at the University of Stellenbosch.**

Supervisors: Mr R. Burger and Prof S. van der Berg

March 2007

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature.....

Date:.....



ACKNOWLEDGEMENTS

To God alone be the glory.

I would like to thank Rulof Burger for his extensive assistance and inputs into this paper. I would like to thank Prof Servaas van der Berg for his perceptive comments during the writing of this paper.

The comments of participants at the DPRU/TIPS Conference on “Accelerated and Shared Growth in South Africa: Determinants, Constraints and Opportunities” are gratefully acknowledged.

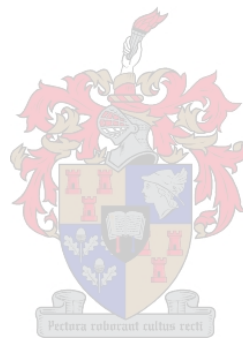


Table of Contents

1	Introduction.....	1
2	Unemployment in South Africa.....	3
3	Data and methodology.....	7
3.1	Methodology.....	7
3.1.1	Pseudo-panels.....	7
3.1.2	Decomposition Analyses – Age, Generational and Year.....	11
3.2	Data description.....	14
3.3	Preliminary Evaluation of Various Estimators on the OHS/LFS synthetic panel....	17
3.3.1	Pooled Ordinary Least Squares (POLS).....	18
3.3.2	Fixed Effects or Least Squares Dummy Variables (LSDV).....	18
3.3.3	Generalised Method of Moments (GMM) Estimation.....	20
3.3.4	Testing for the most appropriate estimator.....	21
3.3.4.1	Developing the testing procedures.....	21
3.3.4.2	Test Results.....	23
4	Decomposition of unemployment rate and its constituent parts.....	30
4.1	Birth cohort panel decompositions – the whole population.....	30
4.1.1	Decomposition of unemployment rate by age, cohort and year.....	31
4.1.2	Decomposition of employment and participation rates by age, cohort and year .	34
4.2	Birth cohort panel decompositions – by population group.....	35
4.2.1	Decomposition of unemployment rate by age, cohort and yea.....	39
4.2.2	Decomposition of employment and participation rates by age, cohort and year .	40
4.2.3	Decompositions with controls.....	46
4.2.3.1	Descriptive analysis.....	46
4.2.3.2	Control Variables.....	49
4.2.3.3	Effect of controls on profile differentials.....	51
4.3	Decompositions by education level.....	58
4.3.1	Decomposition of unemployment rate by age, cohort and year.....	60
4.3.2	Decomposition of employment and participation rates by age, cohort and year .	65
5	Conclusion.....	69

6	Bibliography	71
	Appendix 1: Code to compare GMM and FE Estimators	75
	Appendix 2: Summary of Cohort Sizes.....	80
	Appendix 3: Presentations in 3 Dimensions: Age, Year and Birth Cohort.....	81

List of Figures

Figure 1	Predicted Cross-sectional Unemployment Rates by Age: OHS 1995 and LFS 2005b.....	14
Figure 2	Unemployment rate and Formal unemployment rate, 1995-2005	15
Figure 3	Comparison of FE and GMM estimates: Unemployment rate by cohort and population group with decompositions, 1995-2005	25
Figure 4	Comparison of FE and GMM estimates: Employment rate by cohort and population group with decompositions, 1995-2005	26
Figure 5	Comparison of FE and GMM estimates: Participation rate by cohort and population group with decompositions, 1995-2005	27
Figure 6	Unemployment rate by cohort with decompositions, 1995-2005	33
Figure 7	Employment rate by cohort with decompositions, 1995-2005	37
Figure 8	Participation rate by cohort with decompositions, 1995-2005.....	38
Figure 9	Unemployment rate by birth cohort and population group with decompositions, 1995-2005	43
Figure 10	Employment rate by birth cohort and population group with decompositions, 1995-2005..	44
Figure 11	Participation rate by birth cohort and population group with decompositions, 1995-2005..	45
Figure 12	Educational attainment, by birth year and population group	47
Figure 13	Proportion of over-aged learners, by year and population group.....	47
Figure 14	Proportion of individuals living in rural areas, by age and population group.....	48
Figure 15	Unemployment rate by cohort with decompositions – with controls, 1995-2005	55
Figure 16	Employment rate by cohort with decompositions – with controls, 1995-2005	56
Figure 17	Participation rate by cohort with decompositions – with controls, 1995-2005.....	57
Figure 18	Black unemployment rate by birth and education cohorts, with decompositions, 1995-2005	63
Figure 19	Black employment rate by birth and education cohorts, with decompositions, 1995-2005 .	67
Figure 20	Black participation rate by birth and education cohorts, with decompositions, 1995-2005 .	68
Figure A 1	Unemployment rate by birth cohort and year	81
Figure A 2	Employment rate by birth cohort and year.....	82
Figure A 3	Participation rate by birth cohort and year	82

List of Tables

Table 1	Tests of Coefficient Equality ($H_0: \beta_{GMM} = \beta_{FE}$): Age Effects (P-values).....	28
Table 2	Tests of Coefficient Equality ($H_0: \beta_{GMM} = \beta_{FE}$): Year Effects (P-values).....	29
Table 3	Control variable coefficients	52
Table 4	Cohort sizes: Birth Cohort Panel.....	80
Table 5	Cohort sizes: Birth and Population Group Cohort Panel.....	80
Table 6	Cohort sizes: Birth and Education Level Cohort Panel (Black).....	81

1 Introduction

A new political dispensation in 1994 heralded a period of optimism for many ordinary South Africans, who hoped for freedom and an escape from poverty. Since this transition, however, South Africa has registered steady increases in unemployment, which was already high and widespread at that stage. The new policy environment introduced a mix of legislation which changed the way in which South African society was to be structured: separate development was abandoned, the pillars of Apartheid dismantled, and equitable access to education and jobs was enacted. At the same time, the Reconstruction and Development Programme (RDP), as well as the Growth Employment and Redistribution (GEAR) document addressed, amongst other issues, socioeconomic and labour market disparities. Economic growth was to bolster employment generation. Rising unemployment is, in light of these diverse changes, a source of considerable concern to labour market participants and policymakers alike: the benefits of better understanding the dynamic forces at play are potentially large. Given the many and far-reaching changes referred to above, it is a complex task to disentangle specific reasons for the outcomes realised in the labour market, and more so the manner in which these have interacted to arrive at the status quo.

Since the inception of the October Household Surveys in 1993 and their successors, the Labour Force Surveys, the scope for labour market research has been extended substantially. Comprehensive data in successive periods allows a clearer picture of the state of the labour market to emerge. This data, however, has limitations and restricts the extent to which dynamics can be extracted. The South African literature has been dominated by comparative static analyses, due to the lack of reliable panel data and comparability issues. This strategy, while instructive, omits many underlying features which require tracking the sampled population over time. Such information has eluded analysts, yet the number of surveys now available allows the use of alternative techniques.

In this paper, a synthetic panel approach is adopted, in order to take advantage of the wealth of information which has accumulated in the form of seventeen successive cross-sectional household surveys. South African unemployment is known to have a strong age dimension. By following the mean characteristics of groups of individuals born in the same year from a pre-specified sub-population (birth cohorts), the cohort panel methodology is ideally suited to tease out more information regarding this aspect of unemployment. To fully understand life cycle patterns in the context of a changing economic milieu, it is necessary to trace the flows

of these groups over time, rather than comparing older with younger individuals within the same cross section.

To paint a comprehensive picture, both sides of the market are analysed. “Jobless growth” suggests that labour *demand* has been slow to respond to new economic opportunities. Labour *supply*, however, has added significant pressure to the scenario. Which factors determine why the formal sector does not absorb this emergent activity? Skills mismatches have been emphasised in recent policy discussion, despite large strides in educational attainment. While the labour market has been supplied with larger inflows of better qualified individuals, the goal posts have shifted due to structural economic change. In the process, better educated individuals expect to be more marketable, which fuels labour supply. Yet, the changing demand structure leaves most of these participants without suitable work.

The empirical approach adopted in this paper first seeks to establish variation in both general supply and demand patterns in South Africa by constructing a pure birth cohort panel. It further investigates specific racial features and possible explanations for these, and then proceeds to focus the analysis on educational outcomes to uncover behaviour at various skills thresholds. In so doing, the experiences of different groups can be compared not only with each other, but in different time periods and generations, and over a typical life cycle.

The paper is structured as follows. Section 2 explores some of the pressing questions related to unemployment in South Africa. Section 3 outlines methodological and data issues. These are determined largely by the type of data at our disposal, namely successive cross sections. Due in part to concerns of data quality, this paper restricts itself to studying changes in the formal economy. Potential obstacles related to synthetic panels are identified, and subjected to various methods to find the appropriate tools for continued applied analysis. Section 4 proceeds with a decomposition strategy introduced by Deaton (1997). Unemployment is broken down into cyclical, generational and life cycle components. The paper then investigates the respective roles that changes in the participation and employment rates play in driving unemployment. Racial differentials are examined, after which controls are introduced in an attempt to explain the variation in cohort-specific fixed effects. An educational pseudo-panel delves deeper to trace the experiences of participants at different skill levels. Section 5 concludes the study.

2 Unemployment in South Africa

“Jobless growth” is a phrase which has dominated media, political and academic commentary on the South African economy in recent years. ASGISA, the government’s most recent offering on the economic policy front, has the objective of “shared growth”: this implies that economic growth should indeed translate to the absorption of the large unemployed and inactive work force into the fray of the formal economy. Unemployment is of foremost concern in a developing country such as South Africa and it “demands an explanation” (Kingdon and Knight, 2004: 404). How can objectives of “shared growth” be met, and is policy in the position to offer solutions to these problems?

Jobless growth entails that economic expansion proceeds with capital intensity, as opposed to labour intensity. This mode of economic activity persists despite a large and growing unskilled labour force, which may remain unemployed as a result of higher than market clearing wages. This in turn could be the product of an inflexible labour market, which can be potentially attributed to legislation enacted after the political transition (Burger & Woolard, 2005: 4). While these demand factors pose important questions, the change in supply patterns have proven to be of greater interest. It appears that previously marginalised groups have embarked on a movement into the labour force, which has grown considerably faster than the working age population in recent years: much of this may be attributed to the steady stream of female entrants (Burger & Woolard, 2005; Casale & Posel, 2002; Branson, 2006). While this group has gained more employment, participation has outpaced this progress to contribute to rising unemployment. The economy is therefore unable to accommodate the inflows into the labour force, as the initial stocks of unemployed individuals remain a large obstacle. Branson (2006) shows that the proportion of the population employed has remained stable over the 1995-2004 period; the role of accelerating participation is therefore emphasised in increasing unemployment. Furthermore, black youths have chosen to enter the labour market much earlier than previous generations, which contributes to the participation surge (Branson, 2006). Can some of these trends be explained?

Casale & Posel (2002: 172) hypothesise that higher educational attainment and lower fertility rates among females have induced the exodus from inactivity to labour supply. Lam & Anderson (2001) confirm that at low levels of education, fertility remains high and labour market participation low; once women achieve substantially high levels of education, the choice to enter the labour market predominates. In an environment with increasing educational attainment, this apparent trade-off leans towards the choice for fewer children in

whom more can be invested. This does, however, entail a feed through to labour market outcomes, in particular a concurrent escalation of participation rates, which in turn fuel unemployment. If education is to facilitate a transition to a “better life” it seems ironic that improved levels thereof have had the opposite effect in the labour market. In an economy which is “skills hungry”, the unexpected consequences of education result in disappointment. This phenomenon is termed the “educated youth” hypothesis, whereby better qualified individuals often find themselves unemployed as a result of an emergent over-supply of labour from this group (Leibbrandt & Mlatsheni, 2004: 12). While this usually applies to the study of graduate unemployment, it is possible that this problem surfaces at even lower levels of education, where labour surpluses undermine the value of a certain qualification.

Further clues are to be found in the reward structure prominent in the South African labour market: Keswell & Poswell (2004) find strong convexities in the returns to education, with modest gains to be had until the completion of secondary school; thereafter, prospects with higher education improve substantially. This paints a skewed picture, one which however reflects the skills requirements of the economy. Indeed, Ljungqvist (2000: 125) sketches a hypothetical scenario which resembles that of South Africa very closely: the economy has an abundance of *unskilled* labour, which consequently results in large labour market premia for those who are endowed *with skills*; at the same time – despite high future returns to human capital investments – the initial costs of education may prove too burdensome for the impoverished unskilled, so that this investment is foregone. A vicious cycle ensues, with rich skilled individuals able to capitalise on education, while a large section remains unable to escape what may be viewed as a poverty trap.

Louw et al (2006: 4) highlight a number of reasons why the marginal private benefits of education may be considerably higher for richer families, each of which raises the incentive to continue with such investment: among these, richer households may be offered superior quality of education, as a result of their prominent influence in socially, politically and economically powerful circles. Also, more extensive networking opportunities in the “right places” provide richer individuals with greater chances of finding work. This additional advantage favours more investments in education. For initially poor individuals, the continued investment to progress to “high reward” tertiary education is often an unreachable proposition. This is true in the context of better attainment, where matriculation is indeed a reachable, but perhaps insufficient attainment level. Individuals who are (inadequately) educated may therefore supply excessive labour in response to high expectations of improved

quality of life and higher incomes; this supply is not necessarily absorbed as a result of the even higher skill requirements implicit in employability criteria.

Indeed, it is not uncommon to find that human capital investments actually reduce satisfaction of individuals, as a result of unrealised initial perceptions. Powdthavee (2003:7-8) concludes this for South Africans who attained only Grades 11 and 12. Clark & Oswald (1996: 373-375) show this to be true for all successive levels of education in Britain; it should be noted, however, that these results may have been driven by the economic recession at the time. Studies most often confirm a positive association between education and wellbeing, which reflects the returns inherent to human capital theory (see for instance Graham and Pettinato, 2002). The contrary evidence above suggests that certain sociopolitical circumstances (such as recession, high unemployment or political instability) destabilise this relationship, particularly if expected and realised returns differ markedly as a result of such disturbances. Indeed, Urdal (2006: 612) sketches the potential social problems which might ensue, should educated youths not be absorbed into the labour market: high expectations convert to frustration (instead of returns), which fuels internal conflict and in some cases revolutions.

Furthermore, strong associations between parents' and children's' education entails that it may take a number of generations to effect notable changes towards relevant attainment levels: matriculation, in particular has been identified as a "bottle neck" in South Africa, restraining a satisfactory flow through the education system (Louw et al, 2006: 23). Labour supply accumulates at this point, though it may not all be absorbed. Post-apartheid South Africa's concerted attempts at education investment should strike at this apparent poverty trap. It may be that many individuals are helped to reach considerably beyond their initial educational status; this effort may, however, not be far enough to reach truly high return levels by way of *higher* education.

Should youths achieve higher levels of attainment than their parents, it is possible that some cross the necessary skills threshold to receive the expected rewards in the labour market. Many who start from a low base and achieve only a small increment in human capital may, however, expect higher returns than those offered. These misperceptions could well fuel unemployment amongst younger generations. Rospabé & Mlatsheni (2002) indeed highlight the substantially higher unemployment rates among the youth in South Africa. Youths are prone to misperception as a result of inexperience and underdeveloped social networks; together, these factors prompt unemployment. Human capital theory suggests that education

usually offers labour market gains (Cahuc & Zylberberg, 2004: 64): it endows individuals with marketable skills, but also counters these misperceptions, which may result in a difficult labour market experience. It is clear that individuals need to acquire a critical mass of human capital to be both employable and to establish realistic perceptions of the labour market which they wish to enter. Black disadvantage can be ascribed to poorer qualifications in South Africa (see for instance, Rospabé & Mlatsheni, 2002: 19-21). Successful transition into the labour market therefore requires suitable amounts and types of education.

An extensive literature furthermore documents evidence of scarring (Gregg & Tominey, 2005; Arulampalam et al, 2000; Raum & Roed, 2002; Clark et al, 2001), which links a former state to future labour market outcomes: an initial spell of unemployment reduces future wages and the probability of finding employment. Both result from a loss of human capital in the form of foregone experience, as well as progressively lower reservation wages during a time without work. Should a certain generation, therefore, be particularly affected by unemployment, this state could haunt this group over its life cycle, independent of a change in productive characteristics. Excessive current participation rates (which lead to unemployment) among younger generations may therefore have the ultimate consequence of a trap of low employability well into the future. This becomes a concern when it leads to persistent unemployment.

While pure panel data is most suitable to analyse scarring, the repeated cross sections used below do shed some light on some of these matters. Wittenberg (2002), in particular, interprets age profiles of specific cohorts as “flows” through the labour market. Should individuals of a specific group be unemployed at young ages and experience slow rates of absorption over the life cycle, scarring comes to the fore. A steep downward slope of an age-unemployment profile signifies fast changes, as participants from one age group occupy more jobs than the younger group. The opposite also holds true. A moderate slope is associated with a relatively stagnant position. The methodology suggested below extracts generational effects from age profiles: these can be viewed as the “initial state” experienced by each generation, after which the life cycle experience takes effect. Some caution, however, should be applied in interpreting results in this manner: conclusions are based on the assumption that the age profile is uniform across generations. A more accurate picture demands the interaction of age and generational effects to establish differential degrees of life cycle scarring for each birth cohort. Methods applied here do, however, offer important insights into life cycle and generational dimensions of the labour market

3 Data and methodology

3.1 Methodology

3.1.1 Pseudo-panels

Since the first October Household Surveys (OHS) in 1994, and the introduction of the more consistent Labour Force Surveys (LFS) in 2000, there has been a proliferation of studies that analyse South African unemployment. Differences in questionnaire design and sampling methodology, as well as the inconsistent derivation of labour market measures across surveys, however, complicate direct comparisons of these surveys. Taking a longer-term perspective mitigates some of these concerns. Much of the literature has settled for comparative static analysis, by contrasting circumstances depicted by two satisfactorily spaced cross-sections (for example Borat & Oosthuizen, 2005; Kingdon and Knight, 2005; Dias & Posel, 2006).

It has been shown that South African unemployment has a strong age dimension, and youth unemployment has emerged as one of the most challenging social issues in the South African economy (see Mlatsheni & Rospabe 2002, amongst others). By following Deaton (1985) in constructing a birth cohort panel, this feature of increasing unemployment can be studied from a more dynamic perspective and it allows the exploitation of cross-sectional data between the two endpoints typically chosen. By following the mean characteristics of groups of individuals born in the same year from a pre-specified sub-population (they may or may not be the same individuals in different surveys), it is possible to trace life cycle effects, in addition to the effect of business cycle fluctuations and longer-term trends.

Following groups instead of individuals has both merits and drawbacks. The pseudo-panel approach introduces dynamic views, which are not otherwise possible with the successive cross sections surveyed in South Africa. Variables are aggregated by the chosen cohorts. For instance, a participation dummy is averaged over all the individuals in a cohort (taking Statistics South Africa's sampling weights into account), and consequently represents the estimated participation rate for the specific cohort in each year under consideration. This in itself forms the basis for instructive descriptive analyses to trace, for example, the differences in unemployment paths across racial and gender boundaries (for one application of this method to South African labour market data, see Branson, 2006¹).

¹ The methodology employed by Branson (2006) is largely based in the semi-parametric domain, and focusses on the black subpopulation. The current study depends more explicitly on parametric work and a broader

Coding problems associated with South African household surveys obstruct the identification of the same households in different surveys. This precludes exploitation of the rotating panel design of the Labour Force Surveys. The primary benefit of the *pseudo-panel* methodology partially overcomes this concern: it is not necessary to follow the same individuals over time, but rather analyses investigate the dynamics of “look alike” (to use Deaton’s (1985: 110) terminology). As with pure panel data, it is possible to control for much of the unobserved heterogeneity that plague typical labour market studies. Unlike in a panel dataset, however, attrition is of minor concern, since a set of individuals who meet the grouping criteria appears in each survey, despite the effects of migration, non-response and dissolution of households. A panel of semi-aggregated data allows for a richer analysis than either cross-sectional comparisons or pure time series data can offer: it is possible to lend a dynamic perspective to the investigation, yet maintain a breakdown of the variables under consideration. Deaton (1997: 117) applauds cohort data for providing a meeting point between disaggregated microeconomic information and macroeconomic movements. These properties are exploited in this paper, by “disentangling the generational from life cycle components” (Deaton 1997: 117), which is not possible in either the cross section or time series domains.

The use of pseudo-panels should nevertheless proceed by heeding some of the cautions associated with this data construction. Unfortunately, this methodology does not provide an instrument to study individual transitions from one state to another (such as moving from the discouraged worker status to being an active searcher or finding employment), and is therefore not as informative as a pure panel dataset. Such innovations would enrich the job search and mobility literature. Given the infancy of appropriately constructed, nationally representative individual panels in South Africa, it is not clear that there is any way to address such issues as yet².

Furthermore, the sample average of a variable is not always a good estimator of that cohort’s population mean in each period. Should the mean have a large standard error or be constructed by only a few observations, it is questionable to invoke the law of large numbers, so that sample averages could be plagued by considerable noise. Errors-in-variables are likely to occur and empirical models could suffer from bias and inconsistency. Given that much of

spectrum of the population: these choices demand further care in maintaining empirical regularity, as outlined below.

² The LFS panel has indeed been released recently. This paper nevertheless proceeds with the pseudo-panel approach, as the reliability of the pure panel is yet to be established. Furthermore, pseudo-panels allow the incorporation of the OHS cross sections, which offers a longer-term perspective.

this paper uses these aggregated variables as dependent variables, the nature of the measurement error may not prove to be a serious problem: if it is uncorrelated with any of the independent variables, measurement error is absorbed into the model's error term without consistency losses (Wooldridge, 2002: 71). This, however, may not be true in the current setting. Survey years are fundamental dummy variables modelled in the decompositions proposed below; indeed, the survey quality of OHS 1995 (for instance) has been questioned (Branson, 2006), such that this particular year dummy could be correlated with the measurement error inherent to employment, unemployment and participation. This raises the question whether it is valid to treat a pseudo-panel as if it were a pure panel dataset by applying conventional panel estimators.

If the data generating process is accurately characterised by the unobserved effects model at an individual level, a cohort panel is unlikely to share a similar time-invariant cohort-specific effect (Baltagi, 2005: 193). For a typical individual (i), the following may hold in a specific time period (t):

Equation 1

$$y_{it} = x'_{it}\beta + \mu_i + v_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

where N is the number of individuals surveyed and T the number of periods in the panel.

However, once aggregated by cohort, it is necessary to take cognisance of the fact that different individuals constitute the same cohorts in different periods. The average of the individual fixed effects may therefore not be time-invariant. Hence, the "cohort version" of the fixed effects model would, in its most general form, be represented by:

Equation 2

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\mu}_{ct} + \bar{v}_{ct} \quad c = 1, \dots, C; \quad t = 1, \dots, T$$

where c is the cohort index, t is the time index, C is the number of cohorts, T the number of time periods and the bar notation indicates that the cohort averages of specific quantities are applicable.

This model is only identified if it is assumed that $\bar{\mu}_{ct} = \bar{\mu}_c$ (for $t = 1, \dots, T$), which demands that cohorts be constructed with a satisfactory number of individuals in each group. This precludes analyses based on very different samples, yet drawn from the same cohorts in subsequent periods. Large cohort sizes ensure convergence to the population mean in each

period, so that a common unobserved cohort effect can be successfully isolated. The same holds true for lagged dependent variables: y_{it} is often determined by y_{it-1} ; clearly if the same individuals are not used to construct the means, measurement error could potentially conceal autoregressive relationships (Hsiao, 2003: 284). Since cohort sample averages converge to the population means for a large number of observations per cohort, a tradeoff between additional degrees of freedom (by increasing the *number* of cohorts) against *larger* cohorts (which mitigates errors-in-variables obstacles) emerges.

Deaton's (1985) initial work already proposed an adjusted fixed effects estimator, which scales each cohort by the square root of its constituent size and accommodates the covariance structure of the means. This compensates for cohorts with small sizes and cohort means with large standard errors. Unfortunately, this procedure complicates estimation, and may not result in considerable consistency gains. In practise, applied researchers often choose to ignore measurement error issues: it has been shown that with 100 or more observations within each cohort, bias is minimal and adjustments can be safely ignored (Verbeek and Nijman, 1992). In order for the standard fixed effects estimators to be valid, it is therefore important to choose cohorts with a sufficient number of observations. The applied empirical analysis commences by constructing only a birth cohort panel in section 4.1. Section 4.2 aggregates the data to a higher level, by also grouping according to race, which entails a reduction in cohort sizes. Section 4.3 zooms in further, by considering diverse education cohorts among the black group. This raises the additional point that group sizes may not be identical or even similar across cohorts or time. The analysis of Indians, for example, becomes hazardous due to their small sample size. Inoue (2005) addresses differential group sizes, along with efficiency and inferential concerns, by way of a GMM estimator: the weighting matrix consists of relative cohort sizes, and hence acknowledges information which standard panel estimators do not. This estimator, however, also bases inference on the assumption that $N_{ct} \rightarrow \infty$ ($c = 1, \dots, C$ and $t = 1, \dots, T$), or that all cohort sizes are sufficiently large in each time period. Fixed effects estimators in this setting additionally require sufficient cross-sectional degrees of freedom (*as* $C \rightarrow \infty$). Treating synthetic panels as conventional panel data therefore continues to be conditional on satisfactory cohort sizes. To establish whether final results presented below are sensitive to the estimator chosen, section 3.3 precedes the main empirical analysis with comparisons of the alternatives.

3.1.2 Decomposition Analyses – Age, Generational and Year

One of the benefits of cohort data, as mentioned above, is its ability to discern between life cycle, generational and cyclical macroeconomic components of the dependent variable of interest. Deaton (1997: 123-127) outlines how a simple least squares dummy variable (LSDV) regression can feasibly decompose unemployment, employment and participation into these respective elements. Time dummies are included to capture macroeconomic shocks (but also to account for cross-section specific measurement error). Their coefficient sizes can be compared to the business cycle to assess how responsive the labour market is (in terms of creating and shedding jobs) to fluctuations. Dummies for each birth cohort show how the “fixed effects” of each *generation* differ. These cohort effects capture, inter alia, the impact of *long-term* macroeconomic trends and changes in the average set of productive and demographic characteristics across generations of South Africans, independent of the usual life cycle effects. The third set of dummies (representing age), isolates life cycle effects. These separate out the “stylised facts” of the flows evident in the South African labour market: should participation rates of all cohorts be moderate at young ages and stronger later in the life cycle, it is evident that many individuals enter the labour market between these stages as a result of underlying supply drivers. The gradient of this profile determines the rate of these flows. Age effects are assumed to be independent of generational effects and the changing macroeconomic milieu. This assumption may be strong, as it supposes that each generation exhibits identical flows across the life cycle.

Controls are subsequently included in section 4.2.3, in an attempt to absorb the explanatory power attributable to observable productive and demographic characteristics. After introducing conditional variables, cohort profiles offer a better description of generational (dis)advantage attributable to long-term economic progress. Higher educational attainment could possibly drive increases in employability; not accounting for these changing productive characteristics could lead to the false conclusion that only increases in economic capacity cause better employment rates. Controls therefore attempt to isolate as many of the underlying drivers of labour market outcomes, without classifying these with other unobservables in the fixed effects parameters. The latter could be as diverse as unmeasurable educational quality and structural demand changes.

The usefulness of this technique is to account for possible sources of different types of unemployment. For instance, does youth unemployment arise because young people from every generation have always suffered this consequence in South Africa (life cycle effect)? Or

is youth unemployment the product of increasingly rigid labour markets or a declining quality of education, which erodes the skills base (cohort effects)? It is possible to separate flows from initial states, and to provide some insight into labour market scarring, as referred to above. In answering these questions, decompositions are executed for the unemployment rate, but also for its underlying components. The unemployment rate, u , can be expressed as

Equation 3

$$u = \frac{U}{L} = \frac{L - E}{L} = 1 - \frac{E}{L} = 1 - \frac{E P}{P L} = 1 - \frac{E/P}{L/P} = 1 - \frac{e}{p}$$

where U is the number of unemployed individuals,

E is the number of employed individuals,

L is the labour force, and

P is the population of working age.

The unemployment rate is thus a function of the employment (e) and participation rates (p)³. An increase in unemployment is therefore primarily the result of a decrease in the former, an increase in the latter, or a combination of the two. Therefore, the decomposition of unemployment is augmented by uncovering its composite parts with further estimates executed for the related quantities. This effectively breaks down the above-mentioned contributions to unemployment further into supply-related changes coupled with the current absorptive capacity of the economy.

The practicalities of implementing this decomposition entail solving an identification problem. All year, cohort and age effects should account largely for the dependent variable concerned. Perfect multicollinearity is a problem by definition, since age is a linear function of the current year and the birth year associated with a cohort. It is, however, possible to perform a simple transformation on the year dummies to estimate the equation subject to a zero restriction on the time effects (Deaton, 1997: 126). This makes intuitive sense, since these short-run macroeconomic fluctuations are assumed to average to zero in the long-run. The age dummy representing 65 year olds and the cohort dummy representing those born in

³ Note that even though unemployment is expressed as a proportion of the economically active population, participation and employment rates are expressed as a proportion of the entire working-age population, as a consequence of the above identity. By implication, cohort unemployment rates are constructed using fewer individuals than is the case for employment and participation rates. Hence, unemployment decompositions are more prone to measurement error bias. This introduces a further motivation for studying all three quantities rather than unemployment in isolation.

1930 are omitted to form a reference group (which is present in OHS95, the first period in the sample)⁴, while a set of $T - 2$ new time dummies (omitting the first two years) are created by the following transformation:

Equation 4

$$y_t^* = y_t - [(t - 1)y_2 - (t - 2)y_1] \quad t = 3, \dots, T$$

The time effects for the first and second years can subsequently be recovered by way of the zero restriction.

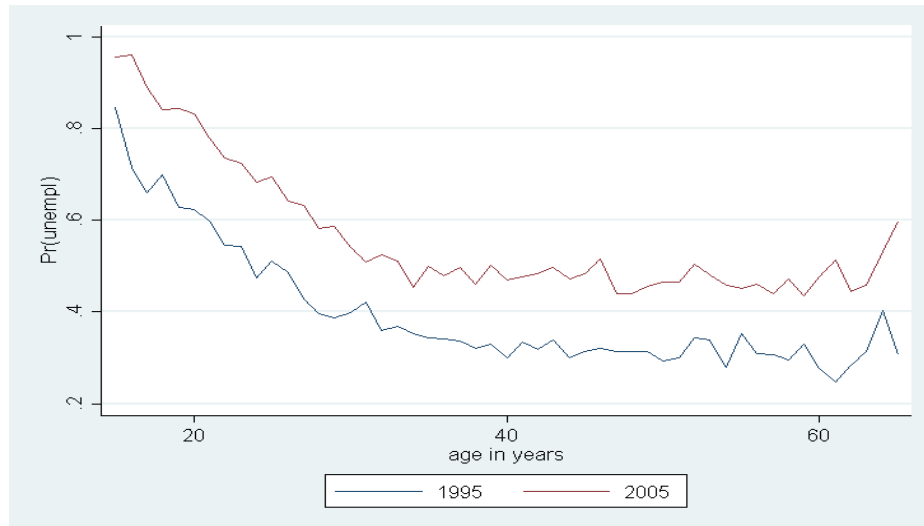
This methodology has been applied to compare racial and gender differentials in South African wages, though only with OHS data (Grün, 2004). Other labour market quantities have been subjected to similar analysis by Branson (2006) with a longer time series. The analysis focussed on outcomes of different genders only within the black population. In this case, the zero restriction was not implemented, thereby only allowing pairwise decompositions, and not the simultaneous isolation of year, life cycle and generational outcomes. This paper attempts to uncover whether unemployment has different drivers for different population groups (increases in participation by specific groups or poorer absorption of different racial categories, which suggests an element of discrimination or skills bias). The time series of eleven years used here allows for a richer picture to unfold.

It is important to note that the decomposition technique employed here ignores interactions between the separate components: the assumption that the age profile of different labour market quantities remains fixed across generations and time is somewhat restrictive. The chosen course of action, however, provides more information than cross section evidence is able to. Figure 1 compares two cross sections (October 1995 and September 2005). Unemployment probits are run on the set of age dummies, after which predicted probabilities of unemployment are obtained for each year. This picture suggests that the age-unemployment profile has remained largely unchanged, but that labour market circumstances have deteriorated uniformly for all age groups. The most credible way to separate age-specific and generational effects is by pseudo-panel decompositions. The decomposition results reported below suggest a flatter profile for younger individuals compared to the cross-section

⁴ This reference group is chosen, since individuals who are 65 years old *and* born in 1930 actually appear in the dataset. Should the youngest birth cohort (born in 1990) and age 65 be chosen, no comparative reference appears in the data. Previous work highlights that results are somewhat sensitive to the choice of an unrealistic reference group. To ensure that conclusions are not affected by this choice, the most sensible known reference group is chosen.

evidence, with more importance accorded to cohort changes in explaining the recent increase in unemployment.

Figure 1 Predicted Cross-sectional Unemployment Rates by Age: OHS 1995 and LFS 2005b



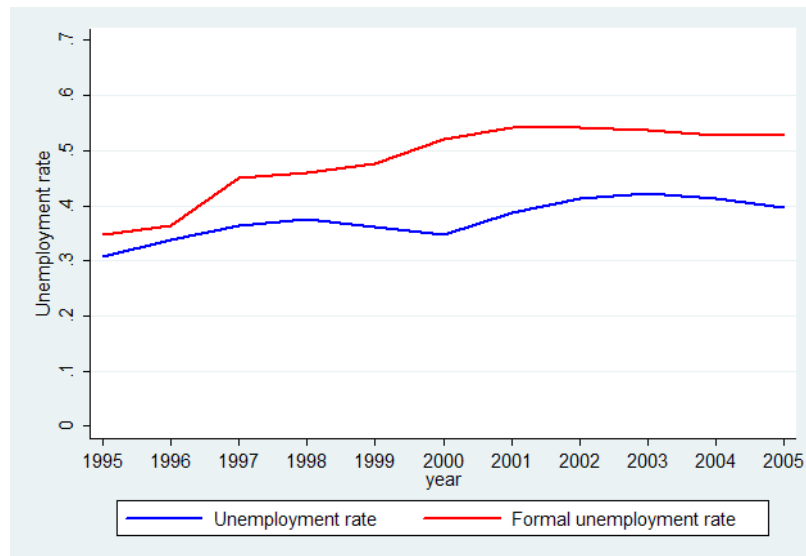
3.2 Data description

This study utilises successive nationally representative South African household surveys conducted between 1995 and 2005, specifically those that focused primarily on labour market issues: these are the 1995 to 1999 annual October Household Surveys, as well as the biannual Labour Force Surveys from 2000 to 2005.

The primary purpose of this study is to analyse changes in the unemployment rate. Kingdon and Knight (2006) – by showing that the non-searching unemployed more closely represent discouraged work-seekers than the voluntarily unemployed – present convincing evidence that the broad (as opposed to the narrow) definition of unemployment more accurately captures the adequacy with which the economy provides employment opportunities to the labour force. For the duration of this paper cohort unemployment and participation rates are calculated using the broad definition of the labour force.

An important consideration in an intertemporal analysis of unemployment is the possibility that inconsistencies in sampling and questionnaire design may distort true trends. This is primarily a concern for the earlier years in the current sample, during which period Statistics South Africa (StatsSA) continually updated questionnaires and improved the sampling design, in order to obtain data that is more reflective of the population. The effect of these changes is particularly evident in the improved capturing of informal economy workers and the large fluctuations registered for agricultural employment (Casale et al, 2004; Burger & Yu, 2006).

Figure 2 Unemployment rate and Formal unemployment rate, 1995-2005



Given the low wages and often unpleasant working conditions faced by informal sector workers (Casale and Posel, 2002), it seems plausible to assume that most labour market participants would consider this an employment option of last resort. Kingdon & Knight (2004: 395) show that formal sector incomes significantly exceed those of the self-employed, and in turn those of the unemployed. The choice of many individuals to stay unemployed rather than enter the informal sector may therefore indicate that the formal job search process is considerably smoother when individuals are not burdened by low income informal business activity. Although the determinants of informal sector employment are of considerable importance in their own right, this study centres around the ability of the formal economy to generate vacancies for a rapidly expanding labour force. If poor formal sector job creation leads more people to resign themselves to a working life in the informal sector, the measure of unemployment should not register this as a decrease in unemployment. For this reason, informal sector workers are not classified with the employed in the calculation of employment and unemployment rates. This furthermore circumvents problems associated with inconsistent capturing of informal employment. It should be noted that since all informal sector workers are now counted as unemployed, the “formal sector unemployment rate” is higher than unemployment rates estimated in the conventional manner. This is evident in Figure 2. The analysis presented below will necessarily be silent on any issues that pertain to the informal economy.

Each of the five OHS's surveyed independently sampled households; the LFS's are based on a rotating panel design, according to which only 20% of sampled households are dropped and

replaced by newly sampled households in each round. StatsSA has not, until recently, released the series of LFS's as a panel dataset. The coding of person identifiers did not facilitate the linkage of individuals across different waves of the LFS (Kingdon and Knight 2005: 18). A number of attempts to re-assemble the panels from the different cross-sections have been executed by academic researchers, but it is too early to judge the reliability of the resulting analyses, and also that of the newly released panel. Devey et al's (2006) work matches 5587 people across five of the LFS waves. However, in the absence of attrition and coding errors, the panel structure should have allowed the comparison of 20% of the 69150 working age individuals (or 13830 observations) that appeared in the February 2002 LFS, which is the first wave they consider. Hence, fewer than 40% of the original observations are apparently recoverable from the LFS's, which implies that studies using the linked data will invariably suffer from a high degree of attrition bias unless this problem can be appropriately addressed.

Until September 2004, the LFS datasets were released with sampling design weights based on the 1996 Census. Subsequent LFS weights were derived from the 2001 Census. In 2005, StatsSA re-weighted all the LFS's that originally used the 1996 Census according to the 2001 Census, in order to aid comparability across surveys. These re-weighted datasets are used in this paper. All the OHS's remain weighted according to the 1996 Census, as this is the closest reference period.

In the creation of the cohort panel, variables are averaged over individuals who share the same birth year. The "birth year" variable is constructed by subtracting individuals' current *age* from the *year* in which the survey was conducted. This variable suffers some measurement error, since all individuals born between the days on which they were surveyed and the 31st of December are assigned the birth year that follows their *actual* year of birth. This could be particularly severe in the March waves of the LFS's, where most individuals are potentially assigned the incorrect birth year. Given that empirical results show no generational discontinuities, this is unlikely to pose a serious problem. This suggests that changes in the labour market are incremental, and that adjacent birth cohorts do not differ vastly from one another. The differences observed between cohorts spaced far apart, however, show how long-term shifts in the labour market emerge. It should be emphasised that pseudo-panels uncover long-term trends most effectively.

The empirical analysis first proceeds without controls to trace only cohort, age and year effects. Subsequently controls are introduced. Geographic heterogeneity is controlled for by

including averaged provincial dummy variables, which represent provincial shares for each cohort. It is also instructive to control for the variation in levels of education across cohorts. Since the employability of cohorts is affected by the distribution of education rather than just its mean, a flexible specification is chosen, that allows for specific effects for each of the different levels of education. The cohort sample means were constructed by averaging over four dummy variables that indicated whether each person had primary, incomplete secondary, complete secondary or any tertiary education. Individuals with NTC I, NTC II qualifications (or who held any certificate or diploma) but have not completed Grade 12 were considered to have incomplete secondary education. Individuals who held an NTC III qualification were counted as having complete secondary education.

An aggregated dummy variable which represents the proportion of over-aged learners in a cohort is also included as a control. Shortly after the political transition, the Department of Education decided to normalise the age profile of learners in schools (Republic of South Africa, 1995, par 33). A part of this process entailed reducing the large numbers of over-age learners (defined in our study to be those older than 19) in school. Steps have been taken to find alternatives for this group in the form of adult education (Republic of South Africa, 1995, par 36). This stance could, however, partially explain surges in labour market participation, as many learners choose not to continue with further education in community learning centres. Furthermore, it is important to consider that the source of the over-age trend can be attributed to high repetition rates: should these individuals choose to exit the schooling system and enter the labour market, incomplete educational attainment reduces the employability of these candidates. In combination, these changes have potentially adverse impacts on unemployment levels in South Africa.

3.3 Preliminary Evaluation of Various Estimators on the OHS/LFS synthetic panel

This section considers which estimators can most adequately elicit the data generating processes (DGP's) of the labour market outcomes, given the specific nature of the data. Empirical illustrations of the relevant propositions are provided in section 3.3.4, from which the most appropriate estimator is chosen to continue the central econometric analyses.

3.3.1 Pooled Ordinary Least Squares (POLS)

The first candidate in any econometric analysis is (Pooled) Ordinary Least Squares (POLS). In this context, the synthetic panel is pooled, under the assumption that both time series and cross section units (birth cohorts in this case) are sampled independently. Essentially, the synthetic panel structure of the data is ignored, and each group in each period is treated as a new observation. Under which circumstances will this strategy deliver consistent estimates? Suppose the DGP following aggregation can be represented as:

Equation 5

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\mu}_c + \bar{v}_{ct} \quad c = 1, \dots, C; \quad t = 1, \dots, T$$

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + u_{ct} \text{ where } u_{ct} = \bar{\mu}_c + \bar{v}_{ct} \text{ is the composite error term}$$

For POLS to be consistent, $E(\bar{x}'_{ct}u_{ct}) = 0$. This, however, demands that $E(\bar{x}'_{ct}\bar{\mu}_c) = E(\bar{x}'_{ct}\bar{v}_{ct}) = 0$ (Wooldridge, 2002: 256). This condition assumes that unobserved cohort effects – which are excluded from the model altogether – do not result in omitted variable bias. This is a stringent assumption in the current context. It is precisely the fact that a number of variables remain omitted in usual cross section analyses, which makes panel methods attractive. As evident below, unobserved cohort effects are explicitly modelled and prove to be heterogeneous across generations. It can be accepted that typically unobserved characteristics account for this variation. How has changing education quality or other socioeconomic features affected labour market outcomes of various generations? It is clear that these omitted factors are of interest, and that they are furthermore likely to be correlated with other unobservables. In particular, age (a regressor) is strongly correlated with birth year (the group-specific variable). Consequently, the POLS approach will not only deliver inconsistent estimates, but also ignore instructive components of the DGP. Inoue (2005: 20) continues to show that the POLS estimator does not converge to a constant, but to a random variable in a pseudo-panel setting. Hence, this estimator stays out of contention from the outset.

3.3.2 Fixed Effects or Least Squares Dummy Variables (LSDV)

Least Squares Dummy Variables regressions (LSDV) extend POLS estimators, simply by adding dummy variables for each group or cohort (bar for a base category) to the list of regressors. This is also embodied in the aforementioned decomposition procedure. The mechanics of this method is mathematically equivalent to the Fixed Effects estimator (Wooldridge, 2002: 272-273), which applies the “within transformation” to each of the

dependent and independent variables. At the individual level (within a pure panel context), this entails demeaning each observation by within-group averages over all time periods. The assumption that the unobserved effects are time invariant implies that these parameters are eliminated from the model post-transformation, while remaining parameters are consistently estimated. As a consequence, the unobserved effects are allowed to be correlated with the other regressors. Under strict exogeneity, this estimator is unbiased and consistent, unlike POLS (Wooldridge, 2002: 268).

Can these properties carry over to a cohort panel? Verbeek and Nijman (1992: 11) maintain that consistency is only achieved if time variation in unobserved cohort effects can be ignored. Should substantial time variation be present, the within transformation does not completely eliminate the unobservables. As a result, the remaining portion is absorbed into the error term. Any correlation of this residual with the other regressors results in inconsistent estimates. Temporal variation can only be ignored if as homogenous groups as possible are sampled in adjacent periods. The rotating panel structure of recent South African labour force survey data introduces some reassurance that those interviewed are almost equally representative of the cohorts they constitute in each survey. This, however, cannot be said of surveys which were conducted prior to this sampling methodology (in particular for the OHS).

The fixed effects (FE) estimator applied to the pseudo-panel follows as:

Equation 6⁵

$$\hat{\beta}_{FE} = (\bar{X}' M' M \bar{X})^{-1} (\bar{X}' M' M \bar{y}) = (\bar{\ddot{X}}' \bar{\ddot{X}})^{-1} (\bar{\ddot{X}}' \bar{\ddot{y}})$$

where \bar{X} is the regressor matrix of cohort averages

and \bar{y} is the vector of independent variable cohort averages

and $M = I_C \otimes \left(I_T - \frac{1}{T} \underline{1}_T \underline{1}'_T \right)$, which is the demeaning matrix

I_s are identity matrices of size s and $\underline{1}_T$ is a column vector of one's of length T

⁵ Note that M is a symmetric idempotent matrix, so that $M'M = MM' = MM = M$, which is often exploited to simplify programming the FE and GMM estimators. It should be noted, however, that this matrix is not necessarily symmetric or idempotent if rows and/or columns are dropped. This may be the case if the panel is unbalanced. The pseudo-panel under consideration is indeed unbalanced, as the youngest birth cohorts only appear in later surveys and older birth cohorts are no longer considered part of working age population in later surveys. To avoid such possibilities, it is necessary to adjust most of Inoue's (2005) estimators and statistics, as that work only considered balanced pseudo-panels.

It is evident that the FE estimator is simply the POLS estimator using time-demeaned regressors ($\bar{\bar{X}}$) and regressands ($\bar{\bar{y}}$)

If this is directly estimated, instead of using the dummy variable approach, it is nevertheless possible to “recover” the unobserved effects’ parameters by the following computation (Wooldridge, 2002: 273):

Equation 7

$$\bar{\mu}_c = \frac{1}{T} \sum_{t=1}^T \bar{y}_{ct} - \left(\frac{1}{T} \sum_{t=1}^T \bar{X}_{ct} \right)' \hat{\beta}_{FE} \quad c = 1, \dots, C$$

This procedure becomes valuable once the GMM estimator is introduced, since it is not as readily connected to the LSDV estimator as the FE estimator is.

Deaton’s (1985) errors in variables fixed effects estimator accounts for potential temporal cohort variation, but is itself only consistent if the total number of observations tends to infinity. A problematic condition, however, is that estimates of the unobserved effect parameters are only consistent if the number of individuals constituting that cohort is large in each time period. Therefore, it is difficult to escape the cohort size predicament, and usual fixed effects estimators, while biased, may prove as satisfactory as error corrections (Verbeek & Nijman, 1992). Consistency of LSDV therefore rests on both the notion that $C \rightarrow \infty$, as well as $N_{ct} \rightarrow \infty$.

3.3.3 Generalised Method of Moments (GMM) Estimation

Inoue (2005) proposes a GMM estimator as an alternative (which, however, still requires that $N_{ct} \rightarrow \infty$ for reliable inference). While its primary purpose is not to account for the “small cohort” bias, the weighting matrix nevertheless exploits information related to relative cohort sizes. Consequently the fixed effects estimator is extended to a static GMM estimator as follows:

Equation 8

$$\begin{aligned} \hat{\beta}_{GMM} &= \left(\bar{X}' M' (M' \hat{\Pi}^{-1} M)^{g_{inv}} M \bar{X} \right)^{-1} \left(\bar{X}' M' (M' \hat{\Pi}^{-1} M)^{g_{inv}} M \bar{y} \right) \\ &= \left(\bar{\bar{X}}' (M' \hat{\Pi}^{-1} M)^{g_{inv}} \bar{\bar{X}} \right)^{-1} \left(\bar{\bar{X}}' (M' \hat{\Pi}^{-1} M)^{g_{inv}} \bar{\bar{y}} \right) \end{aligned}$$

where $\tilde{\Pi}$ is the diagonal weighting matrix with elements $\frac{N_{1t}}{N}, \dots, \frac{N_{2T}}{N}, \dots, \frac{N_{C1}}{N}, \dots, \frac{N_{CT}}{N}$,
 $\tilde{A} = MA$ signifies the demeaned version of A .

The *ginv* operator is the Moore – Penrose generalised inverse for singular matrices.

Other symbols are as before.

The relationship between the GMM estimator and the standard FE estimator is clear from the matrix depiction above. Each cohort observation is weighted by its average relative cohort size over the time series. Smaller cohorts are therefore “weighted up” by the inverse of this factor (once the weighting matrix is inverted). Inoue (2005: 6) points out that the GMM estimator is simply a GLS estimator. GLS *point* estimators (with relative sizes rather than the usual variances as weights) are mechanically equivalent to usual weighted regression analyses which account for sampling design (Lohr, 1999: 361). The covariance structures, however, differ, and consequently interval estimators are not the same. Given the interest here in point estimates for the respective components, the coefficients of the GMM estimator can be interpreted as effects adjusted for “sampling design”, though this design is neither premeditated nor connected to a population sampling frame. Adjustments take into account the importance of each cohort mean in obtaining an aggregate picture.

Inoue (2005: 15-16) shows with Monte-Carlo simulations that (for $N_{ct} \rightarrow \infty$) fixed effects estimators are consistent, but that the distribution of the usual t statistic is distorted in a pseudo-panel context. The GMM estimator is more robust in this regard, and proves to be highly efficient. What are the gains, if the GMM estimator still cannot completely overcome errors in variables? First, coefficients undergo some adjustment by the specification of an appropriate weighting matrix, and secondly inference is improved. This therefore proves to be the “best” available estimator.

Unobserved cohort effects are not explicitly modelled in the GMM estimator, but can be recovered post-estimation, in a similar fashion to the FE estimator.

3.3.4 Testing for the most appropriate estimator

3.3.4.1 Developing the testing procedures

The discussion above highlights that small cohort sizes are not readily overcome by any estimators. It is therefore necessary to choose the “cohort size/observation size” trade-off wisely. In the current context, black cohorts are relatively large, even if disaggregated by further characteristics. Other population groups usually make up smaller proportions of

household surveys. Comparisons across population groups are, however, of interest. This poses the hazard that estimates for any non-black cohorts could be inconsistent and biased. This section attempts to establish to which extent the introduction of GMM estimates improves the outlook for the decomposition procedure. Does weighting by relative cohort size matter? Should coefficients not deviate substantially from fixed effects estimates, the best known course of action is to continue with LSDV estimation. This approach is simple, and should it be revealed as best possible practise, standard econometric software with its built-in conveniences can be exploited. The analyses conducted in this section were programmed in R (R Development Core Team, 2005). The code is available in Appendix 1⁶.

To ensure that conclusions are not swayed by cohort sizes, separate decompositions are estimated for blacks (with sufficiently large cohort sizes of 100 or more, as per Verbeek and Nijman, 1992) and whites (with a larger prevalence of small cohorts) respectively. Profiles are compared both graphically and by formal tests.

The testing procedure is built on the following two asymptotic results from Inoue (2005: 8):

Equation 9

$$Z_{1j} = t_{\beta_{FEj}}^{robust} = \frac{\sqrt{N}(\hat{\beta}_{FEj} - \beta_0)}{\sqrt{\hat{\sigma}_{GMM}^2 [(\bar{X}'M'M\bar{X})^{-1}\bar{X}'M'\hat{\Pi}^{-1}M\bar{X}(\bar{X}'M'M\bar{X})^{-1}]_{jj}}} \sim N(0,1)$$

$$Z_{2j} = t_{\beta_{GMMj}} = \frac{\sqrt{N}(\hat{\beta}_{GMMj} - \beta_0)}{\sqrt{\hat{\sigma}_{GMM}^2 [(\bar{X}'M'(M'\hat{\Pi}^{-1}M)^{g^{inv}}M\bar{X})^{-1}]_{jj}}} \sim N(0,1)$$

where j denotes the position in the coefficient vector

It is now evident that $Z_{1j}^2 \sim \chi^2_1$ and $Z_{2j}^2 \sim \chi^2_1$ by usual statistical properties. Consequently it is possible to construct an F statistic, as the quotient of two chi-square statistics, each scaled by their respective degrees of freedom (one in each case):

Equation 10

$$F_{1j} = \frac{Z_{1j}^2/1}{Z_{2j}^2/1} \sim F(1,1)$$

⁶ I am indebted to Prof Atsushi Inoue for sending me his Matlab code, which was used to trace errors in my own version. Note that the code used here does not correspond entirely with the estimators in Inoue (2005), as unbalanced panels had to be accommodated. This code is also specific to the decomposition scenario associated with the cohort structure present in the OHS/LFS pseudo-panel.

To test $H_0: \hat{\beta}_{FE_i} = \hat{\beta}_{GMM_i}$, take $\beta_0 = 0$ in the respective Z statistics (which usually tests significance of individual regressors). Further reduction allows the comparison of the following test statistic to the critical value of an $F(1,1)$ distribution for the j^{th} coefficient estimated:

Equation 11

$$F_{1j} = \frac{\hat{\beta}_{FE_j}^2 \left[\left(\bar{X}' M' (M' \hat{\Pi}^{-1} M)^{ginv} M \bar{X} \right)^{-1} \right]_{jj}}{\hat{\beta}_{GMM_j}^2 \left[\left(\bar{X}' M' M \bar{X} \right)^{-1} \bar{X}' M' \hat{\Pi}^{-1} M \bar{X} \left(\bar{X}' M' M \bar{X} \right)^{-1} \right]_{jj}} \sim F(1,1)$$

A joint test for each of the age and year vectors can also be executed. Since the sum of n χ^2_1 statistics is distributed χ^2_n :

Equation 12

$$F_2 = \frac{\sum_{j=1}^n Z_{1j}^2 / n}{\sum_{j=1}^n Z_{2j}^2 / n} \sim F(n, n)$$

where n is the length of the vector under consideration (the number of age or year coefficients to be tested jointly).

This simplifies to:

Equation 13

$$F_2 = \frac{\sum_{j=1}^n \left[\hat{\beta}_{FE_j}^2 / \left(\bar{X}' M' M \bar{X} \right)^{-1} \bar{X}' M' \hat{\Pi}^{-1} M \bar{X} \left(\bar{X}' M' M \bar{X} \right)^{-1} \right]_{jj}}{\sum_{j=1}^n \left[\hat{\beta}_{GMM_j}^2 / \left(\bar{X}' M' (M' \hat{\Pi}^{-1} M)^{ginv} M \bar{X} \right)^{-1} \right]_{jj}} \sim F(n, n)$$

It is not necessary to obtain an estimate of $\hat{\sigma}_{GMM}^2$ to conduct these tests. One drawback, however, is that other unmodelled parameters (such as the omitted year effects and the unobservable cohort effects) cannot be tested in this fashion. Inference relating to cohort effects is only possible for LSDV estimates, where time demeaning is not implemented.

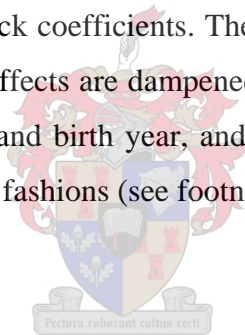
3.3.4.2 Test Results

Figure 3 compares the FE and GMM estimators for unemployment rate decompositions. It is evident that white profiles remain largely unaffected, while substantial adjustments occur for blacks with the alternative estimator. The noise which results from small cohort sizes is not circumvented by use of GMM: this is particularly evident at the endpoints of the white cohort effect profile. Profile shapes are fairly stable, though the black cohort effects show opposite

trends: the downward slope for the FE estimator is improbable, given the consistently contradictory evidence in the rest of this paper. The source of this discrepancy is unknown. It is, however, true that estimates for unobserved effects are not unique and are subject to restrictions chosen (see footnote 7). In this case, the fact that the omitted constant is absorbed by the unobservables (but also affects the coefficients of age), might have a role to play.

Figure 4 illustrates the differences between the fixed effects and GMM estimates for an employment rate decomposition⁷. The shapes of age, cohort and year profiles are not substantially different between the estimates. It is, however, clear that some gaps between the profiles exist. These are wider in the case of the white cohort effects, which highlights a more prominent role of weighting in smaller population groups. Age and year profiles appear to be largely insensitive to the estimator used; however, the cohort profiles display a more distinct separation.

Similar conclusions hold for participation profiles (Figure 5), though in this case weighting results in greater adjustments for black coefficients. The GMM estimator emphasises the age profile somewhat, while the cohort effects are dampened. This shift may also be attributed to the linear dependency between age and birth year, and that the different estimators allocate the “common” element in alternative fashions (see footnote 7).



⁷ These figures appear to be slightly different to those obtained by LSDV estimates in standard econometric software (compare with Figure 9 to Figure 11), yet consistent in direction at each point, bar for unemployment cohort profiles. The reason for this minor discrepancy is that the programmed estimators do not include a constant, and profiles are consequently not adjusted as below. STATA’s built-in fixed effects estimator finds the constant subject to the constraint that the unobserved effects average to zero, which differs from the usual linear regression constant obtained from LSDV estimates (STATA, 2006). Modelled coefficients and standard errors, however, usually remain unaffected by the different approaches, though unobserved effects are all shifted by the same constant. This means that the shapes of profiles should be unchanged, though their absolute magnitudes may differ. A further issue in the decomposition scenario, however, is that age is a linear function of the year of birth, which defines the “unobservable” variable. If cohort effects are shifted by a constant depending on the specific estimation context implemented, then the age profile will also be affected by definition. The manner in which the specific estimation strategy “divides” the effect of the shift parameter between the age and cohort profiles remains unclear. It is therefore important to interpret these graphs in terms of direction, and not magnitude. For these reasons, the profiles presented in this section are not directly comparable to those in the rest of this paper (which were estimated in STATA with constants). The purpose of this section is to evaluate alternative estimators, and not to draw direct conclusions from the decompositions. Models *with* constants are deemed more appropriate for economic interpretation, while programming is simplified by excluding constants here. Statistical comparisons of the estimators’ properties should carry over to cases with constants, despite different coefficient magnitudes .

Figure 3 Comparison of FE and GMM estimates: Unemployment rate by cohort and population group with decompositions, 1995-2005

Figure 3.1 Unemployment rate by birth cohort and age

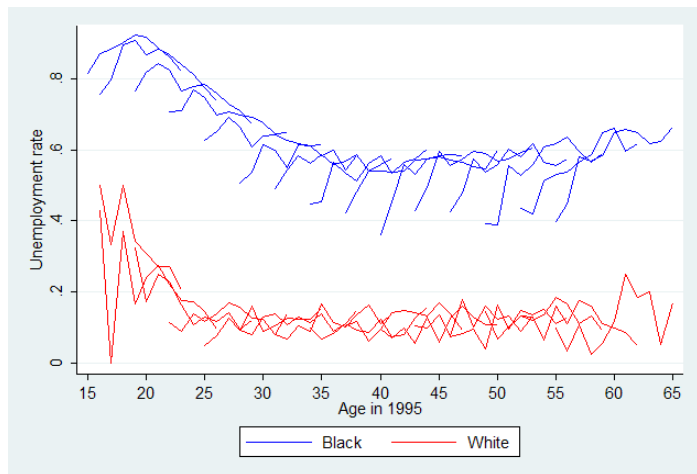


Figure 3.3 Unemployment rate cohort effects

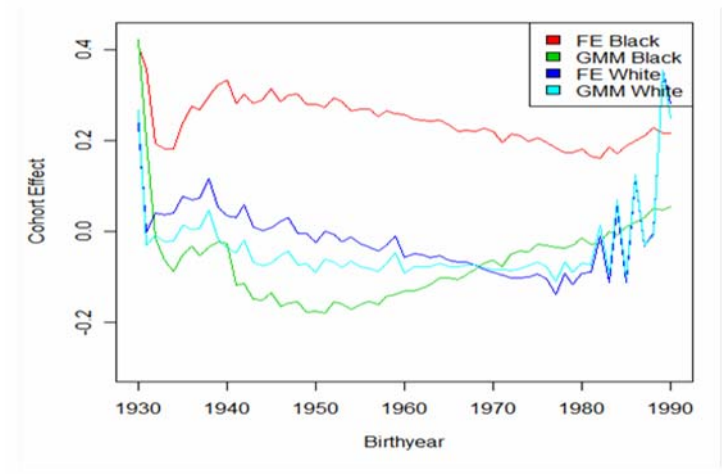


Figure 3.2 Unemployment rate age effects

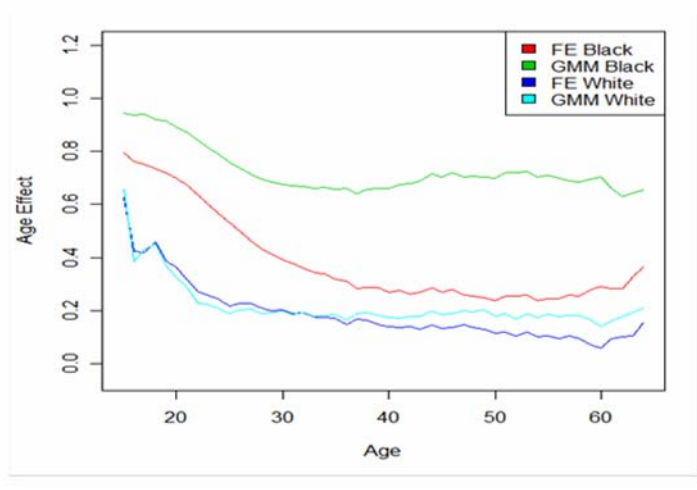


Figure 3.4 Unemployment rate year effects

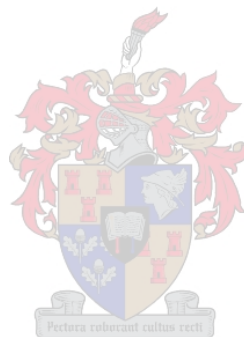
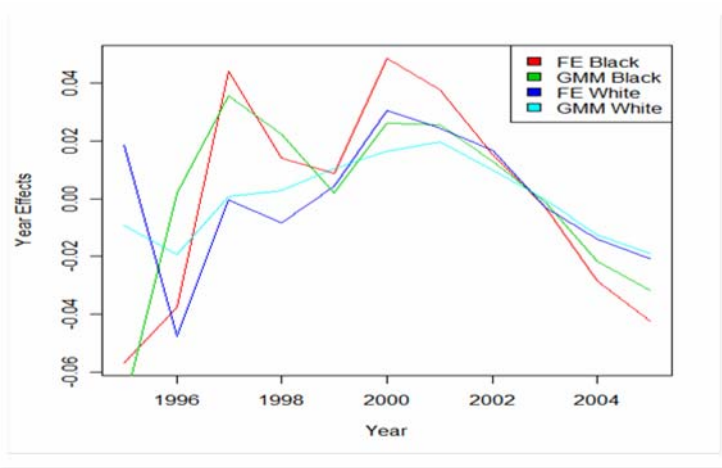


Figure 4 Comparison of FE and GMM estimates: Employment rate by cohort and population group with decompositions, 1995-2005

Figure 4.1 Employment rate by birth cohort and age

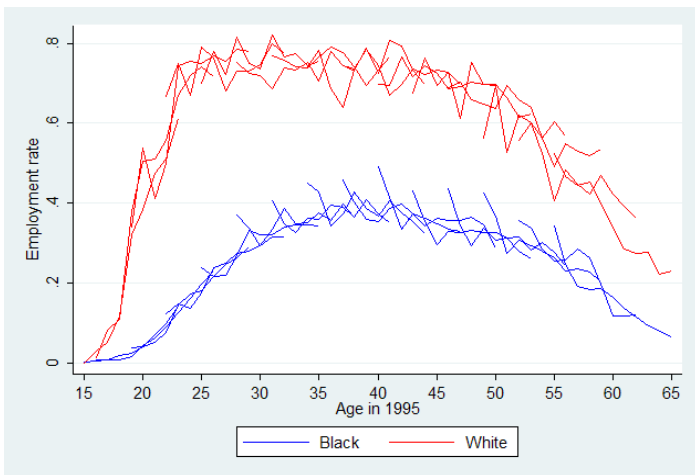


Figure 4.3 Employment rate cohort effects

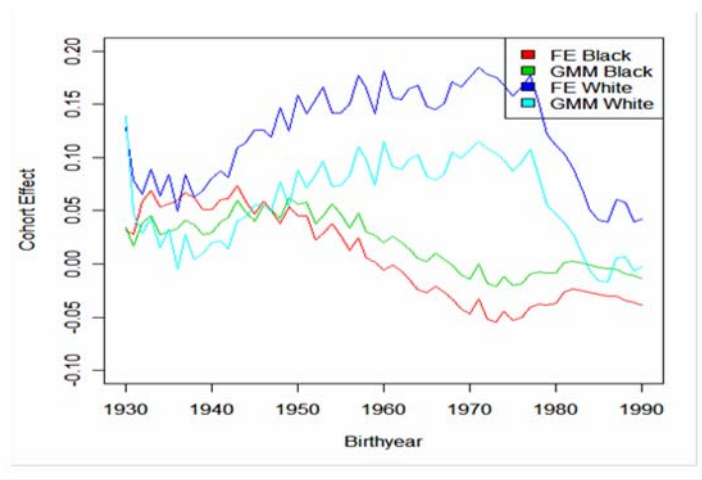


Figure 4.2 Employment rate age effects

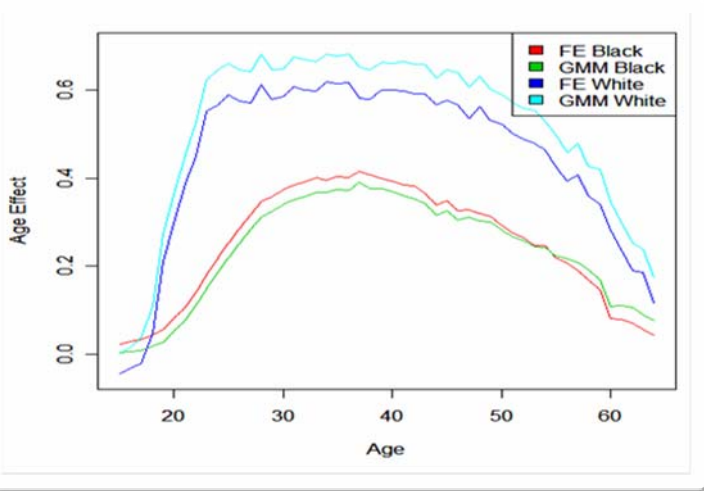


Figure 4.4 Employment rate year effects

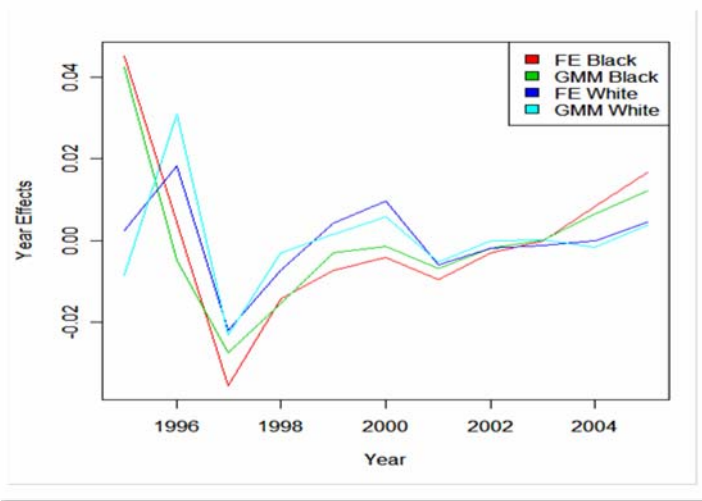


Figure 5 Comparison of FE and GMM estimates: Participation rate by cohort and population group with decompositions, 1995-2005

Figure 5.1 Participation rate by birth cohort and age

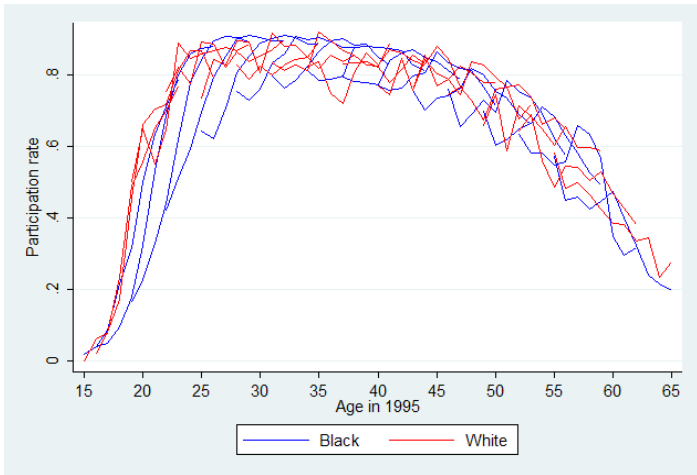


Figure 5.3 Participation rate cohort effects

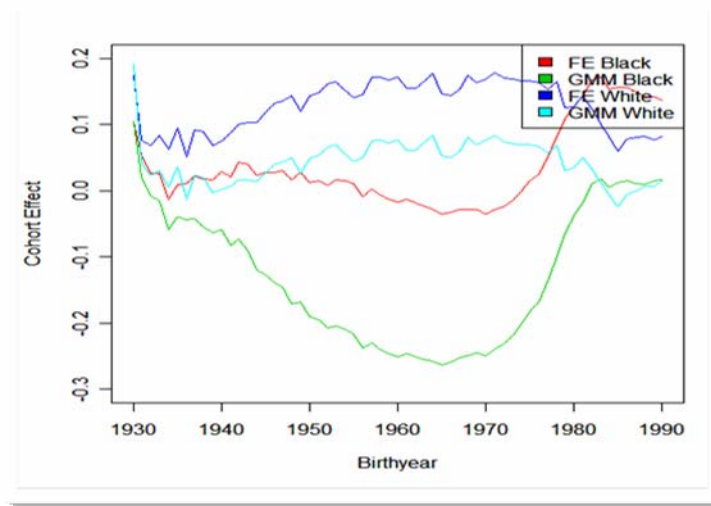


Figure 5.2 Participation rate age effects

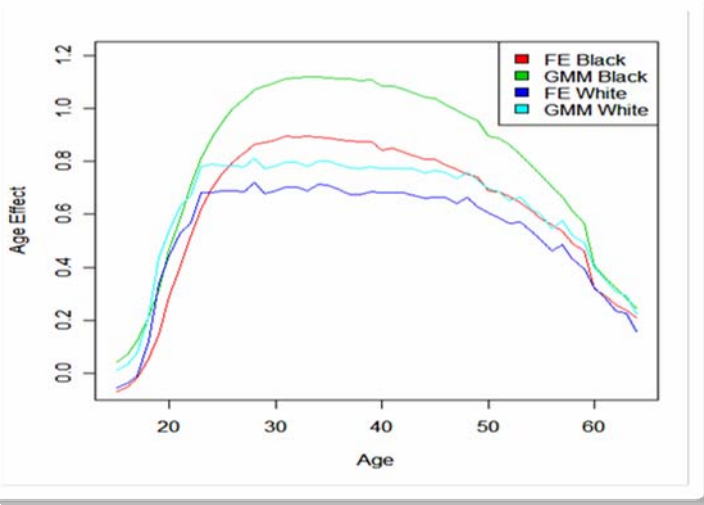


Figure 5.4 Participation rate year effects

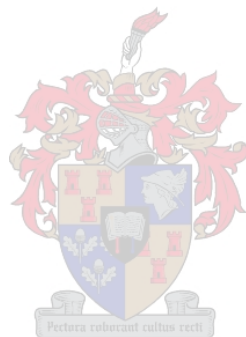
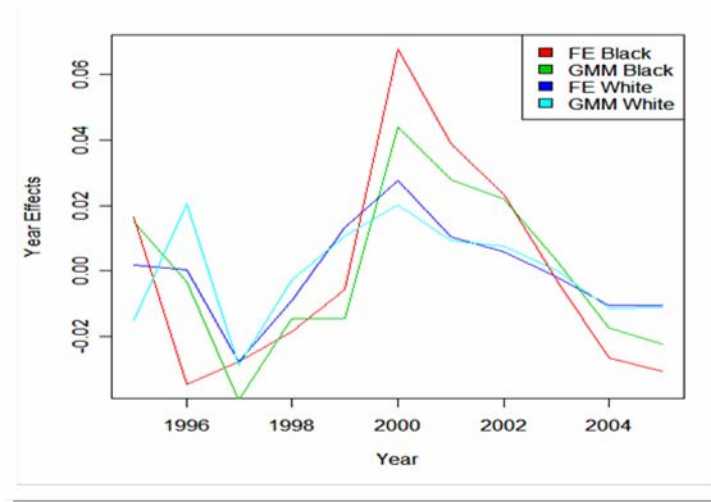


Table 1 Tests of Coefficient Equality ($H_0: \beta_{GMM} = \beta_{FE}$): Age Effects (P-values)

	Black			White		
	Unemployment	Employment	Participation	Unemployment	Employment	Participation
Age15	0.665	0.086 *	0.430	0.594	0.005 ***	0.150
Age16	0.667	0.145	0.691	0.552	0.375	0.537
Age17	0.689	0.210	0.940	0.626	0.730	0.943
Age18	0.731	0.340	0.887	0.700	0.806	0.752
Age19	0.753	0.401	0.806	0.742	0.670	0.668
Age20	0.785	0.491	0.754	0.747	0.652	0.651
Age21	0.802	0.531	0.730	0.761	0.642	0.647
Age22	0.810	0.549	0.709	0.734	0.632	0.635
Age23	0.812	0.556	0.693	0.737	0.615	0.619
Age24	0.808	0.556	0.676	0.727	0.608	0.612
Age25	0.802	0.559	0.670	0.718	0.600	0.605
Age26	0.797	0.558	0.660	0.714	0.598	0.601
Age27	0.803	0.563	0.661	0.715	0.597	0.601
Age28	0.808	0.561	0.658	0.714	0.597	0.601
Age29	0.811	0.563	0.658	0.719	0.594	0.601
Age30	0.818	0.565	0.658	0.731	0.594	0.601
Age31	0.814	0.557	0.649	0.715	0.593	0.599
Age32	0.826	0.564	0.658	0.724	0.597	0.602
Age33	0.825	0.556	0.648	0.715	0.591	0.596
Age34	0.826	0.559	0.649	0.718	0.588	0.594
Age35	0.835	0.562	0.653	0.730	0.590	0.597
Age36	0.836	0.555	0.648	0.734	0.589	0.598
Age37	0.844	0.560	0.650	0.723	0.595	0.601
Age38	0.847	0.556	0.650	0.733	0.592	0.600
Age39	0.846	0.562	0.651	0.738	0.588	0.597
Age40	0.856	0.563	0.656	0.741	0.588	0.596
Age41	0.850	0.554	0.648	0.731	0.587	0.594
Age42	0.862	0.560	0.658	0.740	0.594	0.600
Age43	0.858	0.559	0.654	0.754	0.592	0.601
Age44	0.852	0.552	0.651	0.743	0.588	0.596
Age45	0.861	0.561	0.655	0.756	0.597	0.605
Age46	0.856	0.552	0.648	0.742	0.597	0.602
Age47	0.862	0.562	0.652	0.741	0.607	0.611
Age48	0.866	0.561	0.652	0.749	0.596	0.601
Age49	0.868	0.565	0.653	0.763	0.602	0.610
Age50	0.873	0.563	0.654	0.764	0.603	0.608
Age51	0.865	0.562	0.650	0.762	0.602	0.608
Age52	0.868	0.576	0.659	0.769	0.605	0.609
Age53	0.865	0.578	0.656	0.756	0.612	0.614
Age54	0.869	0.573	0.650	0.778	0.605	0.611
Age55	0.869	0.589	0.657	0.777	0.620	0.625

	Black			White		
	Unemployment	Employment	Participation	Unemployment	Employment	Participation
Age56	0.863	0.588	0.644	0.796	0.624	0.629
Age57	0.855	0.608	0.644	0.778	0.627	0.629
Age58	0.852	0.618	0.646	0.798	0.633	0.636
Age59	0.844	0.634	0.649	0.829	0.654	0.657
Age60	0.835	0.671	0.661	0.828	0.649	0.652
Age61	0.823	0.677	0.649	0.755	0.655	0.652
Age62	0.824	0.710	0.658	0.759	0.668	0.663
Age63	0.790	0.715	0.639	0.765	0.654	0.655
Age64	0.772	0.743	0.640	0.705	0.697	0.679
Joint	1.000	0.947	0.999	1.000	0.977	0.982

P-values reported for coefficients based on an F(1,1) statistic

P-values for the joint test based on an F(50,50) statistic

*Reject H_0 at 10% level of significance

**Reject H_0 at 5% level of significance

***Reject H_0 at 1% level of significance

Table 2 Tests of Coefficient Equality ($H_0: \beta_{GMM} = \beta_{FE}$): Year Effects (P-values)

	Black			White		
	Unemployment	Employment	Participation	Unemployment	Employment	Participation
Year1997	0.560	0.468	0.658	0.772	0.524	0.519
Year1998	0.750	0.572	0.477	0.305	0.254	0.182
Year1999	0.224	0.290	0.800	0.835	0.235	0.437
Year2000	0.394	0.256	0.414	0.403	0.357	0.410
Year2001	0.475	0.446	0.446	0.562	0.471	0.471
Year2002	0.546	0.404	0.527	0.509	0.056 *	0.584
Year2003	0.365	0.454	0.544	0.101	0.028 **	0.087 *
Year2004	0.521	0.466	0.412	0.626	0.955	0.526
Year2005	0.513	0.448	0.447	0.614	0.446	0.524
Joint	0.561	0.515	1.000	0.544	0.569	0.630

P-values reported for coefficients based on an F(1,1) statistic

P-values for the joint test based on an F(9,9) statistic

*Reject H_0 at 10% level of significance

**Reject H_0 at 5% level of significance

***Reject H_0 at 1% level of significance

Table 1 and Table 2 display p-values associated with F statistics of both individual coefficients, as well as the relevant vectors, as outlined above. The hypothesis of individual coefficient equality is rejected only in isolated cases, most of which are found in the white group. Age and year effects as collective profiles do not differ significantly at conventional levels from each other. These hypotheses remain untested for cohort effects, as they are not

explicitly modelled in this setup and do not possess standard errors⁸: it may be that different conclusions prevail in this case. The tests, in conjunction with the relatively unaltered profile shapes, show that conclusions may not be adversely affected by simply continuing with standard estimators. The adjusted estimator, while accounting in a sense for “pseudo-panel design”, can in no way overcome the binding constraint introduced by noisy variables. As such, realised gains are limited, and the rest of this paper continues with a simple LSDV procedure.

4 Decomposition of unemployment rate and its constituent parts

4.1 Birth cohort panel decompositions – the whole population

The empirical analysis commences by grouping all individuals born in the same year into cohorts, and applying the decomposition technique suggested by Deaton (1997) to the cohort unemployment rates. This choice of aggregation delivers 561 cohorts, with an average of 1917 sampled *working-age* individuals per cohort, which is applicable to employment and participation analyses (see Table 4). The largest of these consists of 5277 and the smallest of 289 observations. Twenty-two cohorts, however, contain less than 100 sampled labour force *participants* – the acceptable threshold required to ignore sampling errors (as in section 3.1) – which is relevant for the calculation of cohort unemployment rates. Since the average number of labour force participants per cohort is 1205, only a small number of cohorts will suffer non-negligible sampling errors. Therefore the inconsistency that arises from using conventional fixed effects estimators is unlikely to play an important role in the results presented below.

In section 4.1.1 cohort unemployment rates are calculated by averaging the unemployment dummy over all individuals in a birth cohort. This variable is regressed on a set of age dummies representing the ages 15 to 64 (age 65 is chosen as the reference), on the birth cohort dummies, which signify being born in the years 1931 to 1990 (1930 is the reference birth year) and on the nine transformed year dummies. In section 4.1.2 the same method is applied to the cohort participation and employment rates.

⁸ FE and GMM estimators remove these parameters as a result of the time-demeaning “within transformation”. They therefore do not appear in the model, and can only be recovered, as outlined above. Consequently no inference is possible. The LSDV estimator, while equivalent to the FE estimator, includes the fixed effects as modelled variables; as a result standard errors are available, which enables inference.

4.1.1 Decomposition of unemployment rate by age, cohort and year

Figure 6 shows the decomposition of the unemployment rate into age, cohort and year effects. Figure 6.1 depicts the separate unemployment rates experienced by every third birth cohort over their relevant lifecycles – showing each cohort produces a cluttered graph. For example, the leftmost line represents the unemployment rates for the birth cohort aged 15 in 1995 (or those born in 1980) for each of the years from 1995 and 2005. This curve is positioned above that of the next youngest cohort until the age of 23, where the two curves intersect: this implies that the younger cohort faced higher unemployment than individuals born three years earlier at the same ages. Since most of the lines appear above those directly to their right, individuals born more recently generally experience higher unemployment rates than older birth cohorts. The only reason for an overlap in the curves is that unemployment rates for most cohorts show a decline in the last few years of the sample. It is possible to express the same information in three dimensions, as in Figure A 1.

The year effects (in Figure 6.4) represent the impact of cyclical macroeconomic events on the unemployment rate. It should be noted, however, that period-specific sampling peculiarities add some noise to these components. Since the year effects sum to zero by design, they capture the business cycle variation in the unemployment rate. Between 1995 and 1997 a steep increase in cyclical unemployment can be observed, followed by a brief downturn until 1999, before reaching a turning point in 2000. After this, the unemployment rate showed a steady decline. The magnitude of the increase in cyclical unemployment between 1995 and 1997 is primarily driven by an implausibly large decrease in agricultural employment reported in the OHS's over this period. The South African Reserve Bank identified the third quarter of 1999 as the start of an upswing in the South African economy (SARB, 2006: S159). The cyclical variation in unemployment extracted from household surveys therefore shows some correspondence to the business cycle obtained from macroeconomic data. The cycle derived here, however, appears to lag the official cycle by approximately one year. It is not clear whether this is an issue of comparability, or whether a systematic lagged relationship between the overall business cycle and variations in the labour market exists. Should the latter consideration be true, it is probably because employers only transmit gains from increased production turnover to job creation with some delay.

The age profile of unemployment (Figure 6.2) differs markedly from that presented in Figure 1: this demonstrates the value of using cohort panels rather than cross-sectional analyses. Age appears to be a less decisive determinant of unemployment between the ages of 15 and 40.

Unemployment, however, increases rapidly amongst labour force participants of older ages. An initial experience with unemployment may therefore be exacerbated by pure life cycle effects.

The birth cohort graph (Figure 6.3) reveals that generational effects are more decisive factors underlying the increase in unemployment (as judged by the size of the cohort coefficients, labelled on the y-axis). In many respects, the cohort effects represent the most important of the three composite parts, as it is a reflection of the longer-term trend in the economy. Changes in birth cohort effects can result from structural changes at the macro-economic level, shifts in the preferences of individuals or differences in the productive characteristics (observed or unobserved) across generations. In this case, cohort effects show a deterioration of the labour market prospects of younger generations, which is of obvious concern for new entrants. The cohort effects combined with the age profile explains the U-shaped age-unemployment profile observed in cross-sections: the high unemployment experienced by the young is attributable to the disadvantage they face as a result of entering the labour market in a period of higher unemployment, rather than the inherent fact that they are young. In contrast, the higher unemployment rates amongst older individuals are explained by their age and occur despite profiting from the lower unemployment-cohort effect. The pseudo-panel offers a richer analysis of these effects than cross-sectional data.

It is important to emphasise that the decomposition technique used here does not allow causal inferences on the determinants of unemployment. It does not, for instance, reveal *why* age is positively correlated with unemployment beyond 40. It merely serves to discern which part of unemployment can be ascribed to age, time of labour market entry or period-specific phenomena. These factors are exogenous, and no action by individuals or groups can alter these outcomes: the assumption here is that the entire labour market of a certain age group, in a certain time period, which entered at the same time, are subject to the same conditions, which they cannot individually or collectively alter. In section 4.2.3, the empirical investigation probes further, by controlling for a set of productive and demographic cohort characteristics. The goal is to establish whether controls change profiles, thereby explaining portions of their variation. Individuals have some control over these factors (for instance the choice of higher educational attainment), and profiles which remain after conditioning on other characteristics therefore isolate “fixed”, unchangeable attributes in a purer fashion. It is only in that section that causality can be attached to coefficients.

Figure 6 Unemployment rate by cohort with decompositions, 1995-2005

Figure 6.1 Unemployment rate by birth cohort and age

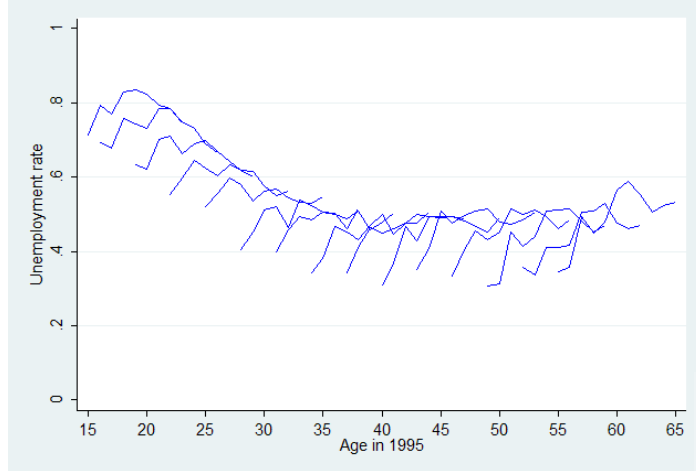


Figure 6.3 Unemployment rate birth cohort effects

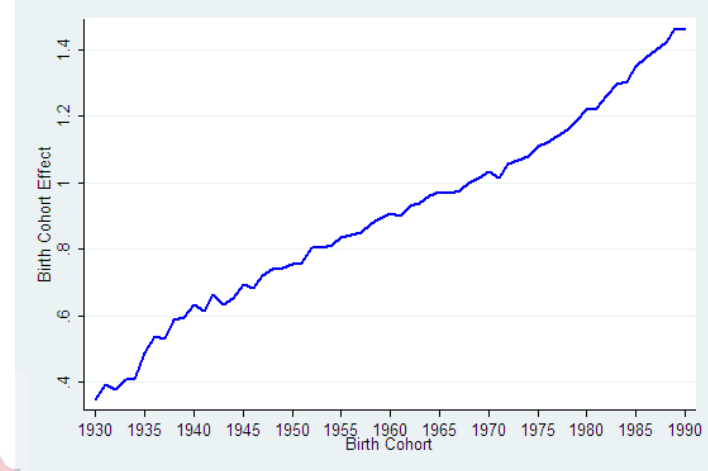


Figure 6.2 Unemployment rate age effects

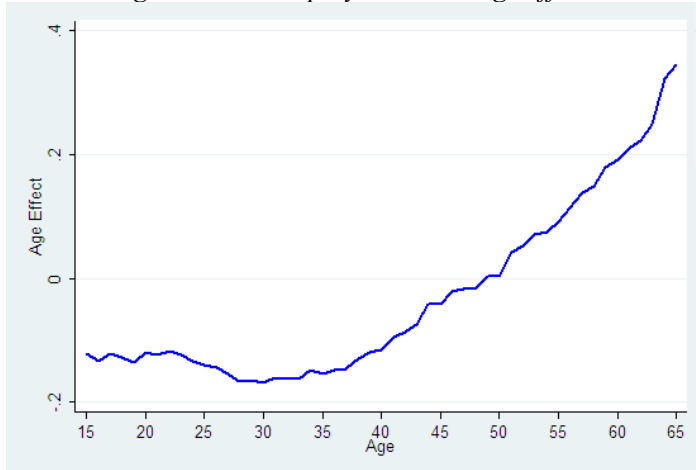
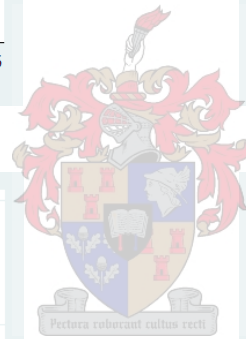
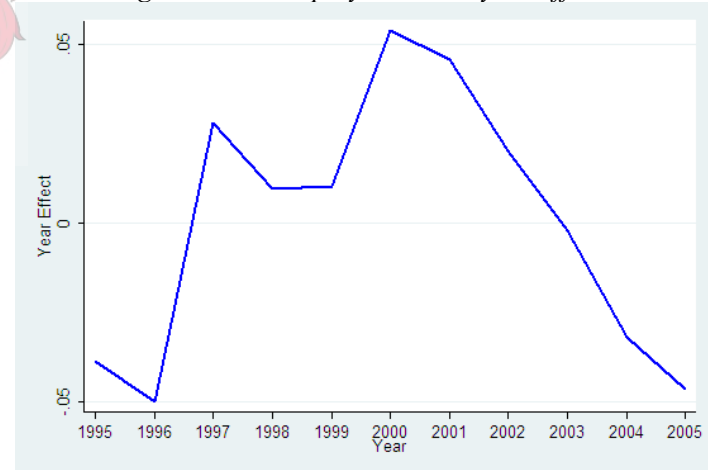


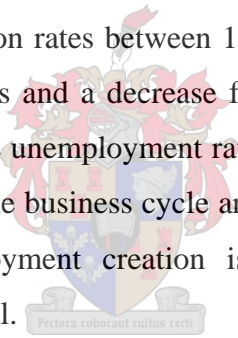
Figure 6.4 Unemployment rate year effects



4.1.2 Decomposition of employment and participation rates by age, cohort and year

In section 3.1 the relationship between the unemployment, and the employment and participation rates was made clear. In order to distinguish between the effects of these composite factors which drive unemployment, the same decomposition technique is applied to the employment and participation rates separately below. The raw cohort employment and participation rates are presented by age and birth year in Figure 7.1 and Figure 8.1 (and by birth year and year in Figure A 2 and Figure A 3).

Figure 7.4 indicates that the large increase in year-specific unemployment between 1995 and 1997 was driven by the rapid decrease in the cyclical component of the employment rate. The source of these changes can be largely attributed to inconsistent sampling methodology, as referred to in section 3.2: over this period the OHS's register an improbably large decrease in agricultural employment. The increase in cyclical unemployment over this period occurred despite a modest decline in the cyclical component of labour force participation (Figure 8.4). Between 1997 and 2005, the employment rate year effects showed a steady resurgence, which dominated the increase in participation rates between 1997 and 1999, but not in 2000. After 2000, stable employment year effects and a decrease for participation combined to cause a drop in the cyclical component of the unemployment rate. The relationships between cyclical employment and participation with the business cycle are not as clear as with unemployment. It is, however, visible that employment creation is procyclical from 2001, and that participation appears to be anticyclical.



The age profiles (Figure 7.2 and Figure 8.2) reveal that both employment and participation are characterised by an inverted U-curve over the life cycle. The similarity of the shape and gradient of these effects for participation and employment below the age of 40 explains why age does not appear to play an important role in determining unemployment: increases in participation are matched by increases in employment, subduing each other's effect on the unemployment rate. After the age of 40, the employment rate starts to drop rapidly, whereas the participation rate is marked by a steadier decline. This means that older labour force participants are "losing jobs" faster than they are leaving the market, which exerts a positive net effect on the unemployment rate.

Birth cohort effects for employment and participation rates are characterised by higher participation rates and lower employment rates for younger generations, both of which imply that unemployment intensifies for this group. The participation increase is roughly linear, but

exhibits a disproportionate surge for those born between about 1975 and 1985. This era is identified by Louw et al (2006: 14) as a time of considerable strides in educational attainment among all population groups, despite persistent differences between races. The controls employed below attempt to verify whether such a connection exists.

From Figure 7.3 and Figure 8.3 it is unclear whether participation cohort effects show more pronounced generational movements than those of employment. This, and the non-linear manner in which these two factors combine to determine unemployment, makes it difficult to gauge the relative contribution of each of these effects by simple coefficient comparisons. For this reason, marginal effects derived from Equation 3 provide a more accurate depiction of each component's relative importance. An increase in the participation rate raises the unemployment rate by approximately $\frac{\delta u}{\delta p} \Delta p = \frac{e}{p^2} \Delta p$ (where e and p denote the employment and participation rates respectively), whereas a decrease in the employment rate increases the unemployment rate by $\frac{\delta u}{\delta e} \Delta e = \frac{1}{p} \Delta e$. At full employment (where $e=p$), the effect of these changes is equivalent. If, however, an incrementally small labour market surplus arises, unemployment rates respond more sharply to a decrease in the employment rate than to an increase in the participation rate. Evaluating these equations at the sample means of e and p (39% and 68%), demonstrates that a change in the employment rate affects the unemployment rate with a magnitude about 60% larger than a similar change in the participation rate would prompt (in percentage points). At the average participation rate, the observed decrease in employment faced by the different birth cohorts would have led to a 19 percentage point differential between the unemployment rates for the youngest and oldest birth cohorts (abstracting from age and year effects), whereas the observed increase in participation rates would have (at the mean employment rate) increased the unemployment rate by 80 percentage points. Therefore, approximately 81% of the increase in unemployment rates faced by the youngest generation was caused by their increased labour force participation rates, whereas the lower employment rate contributed the remaining 19%.

4.2 Birth cohort panel decompositions – by population group

In this section cohorts are further disaggregated by population group to trace differences between the various segments of the labour market. To lend focus to the analysis, and to avoid problems associated with small cohort sizes, the emphasis falls on the comparison of the black and white population groups only. This scenario renders 1122 cohorts, with an average of 1462 working-age observations for black and 167 working-age observations for white

cohorts (see Table 5). The largest cohort consists of 4336 and the smallest of 38 working age observations. For the working-age population, no black cohorts contain fewer than 100 individuals, though for white cohorts this number rises to 119. This deems all black employment and participation analyses consistent, though white estimates pose a somewhat greater risk. As above, for the unemployment decomposition, cohort sizes depend on the number of active labour market participants. Three-hundred such cohorts are constructed from less than 100 labour market participants, 264 of which are from the white population group. The risk of inconsistency arising from measurement error increases substantially when moving to a more disaggregated panel. This is particularly true for modelling white labour market outcomes.

In section 4.1 our interest lay only with the shapes and relative contributions of the different components of unemployment (as well as employment and participation). In comparing the age and birth cohort profiles of blacks and whites, interest now also lies in the level of these curves: therefore the graphs in Figure 9 to Figure 11 were plotted with adjustments to accommodate the effect of the constants from the decomposition regressions. The fact that the year effects are restricted to sum to zero means that the absolute levels of these curves carry no meaning: hence, year effect profiles remain unadjusted.

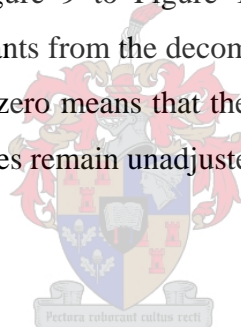


Figure 7 Employment rate by cohort with decompositions, 1995-2005

Figure 7.1 Employment rate by birth cohort and age

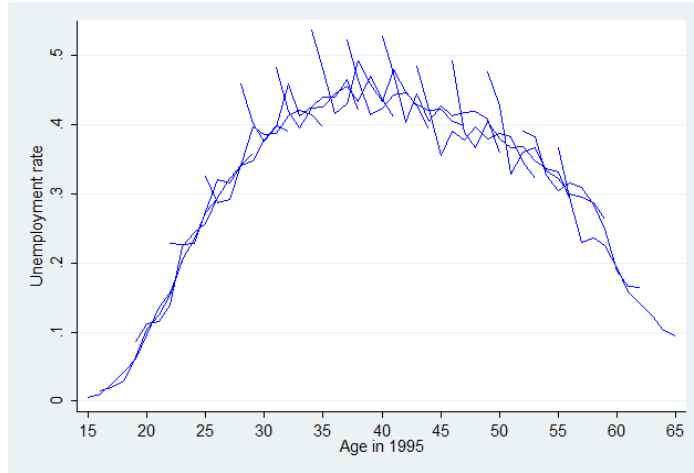


Figure 7.3 Employment rate birth cohort effects

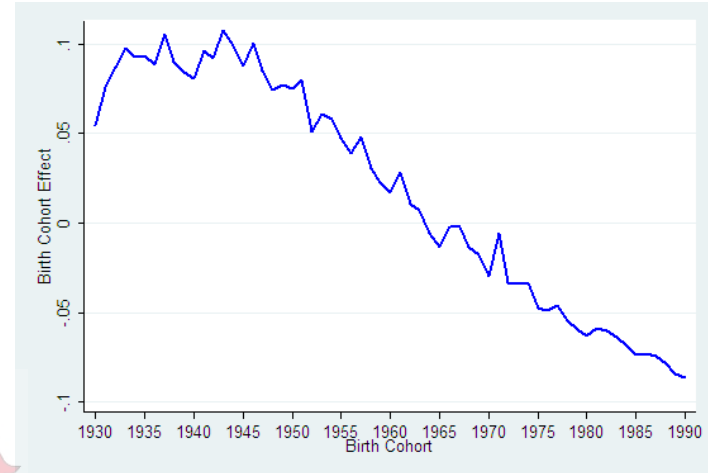


Figure 7.2 Employment rate age effects

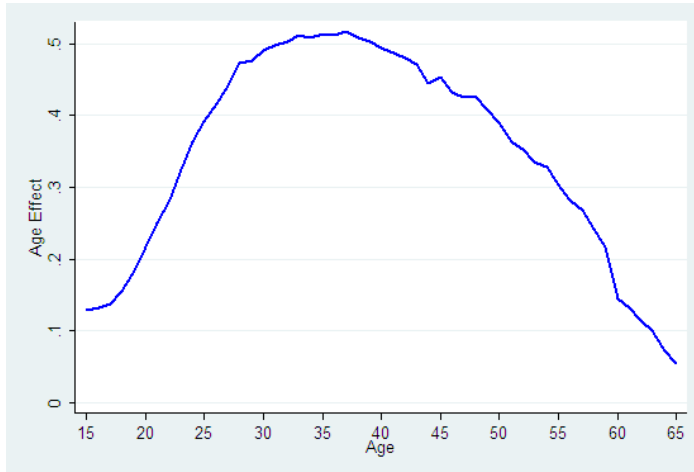


Figure 7.4 Employment rate year effects

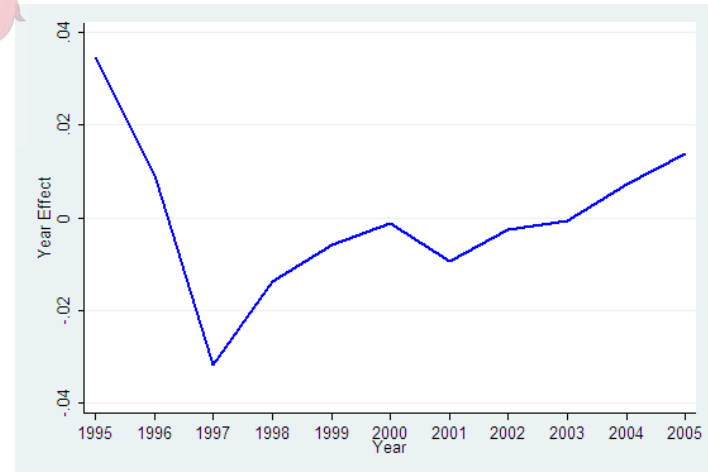


Figure 8 Participation rate by cohort with decompositions, 1995-2005

Figure 8.1 Participation rate by birth cohort and age

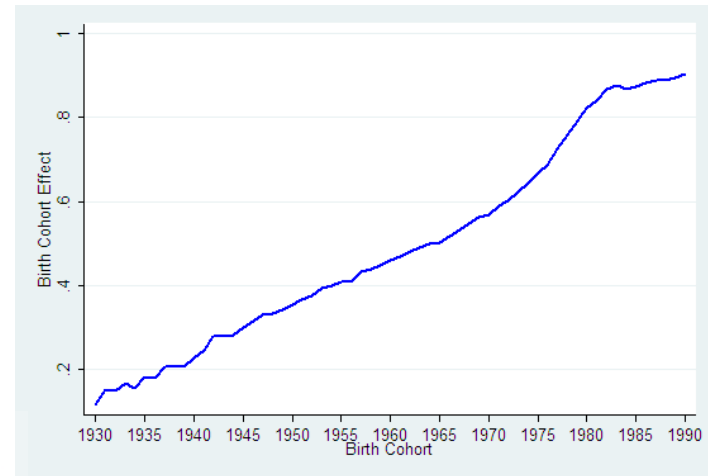
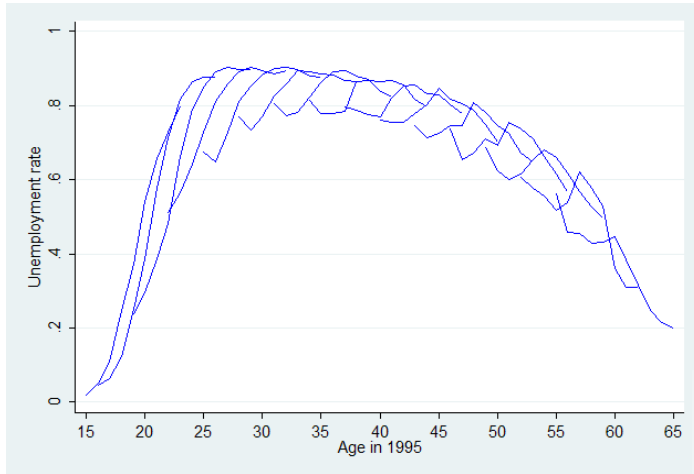


Figure 8.2 Participation rate age effects

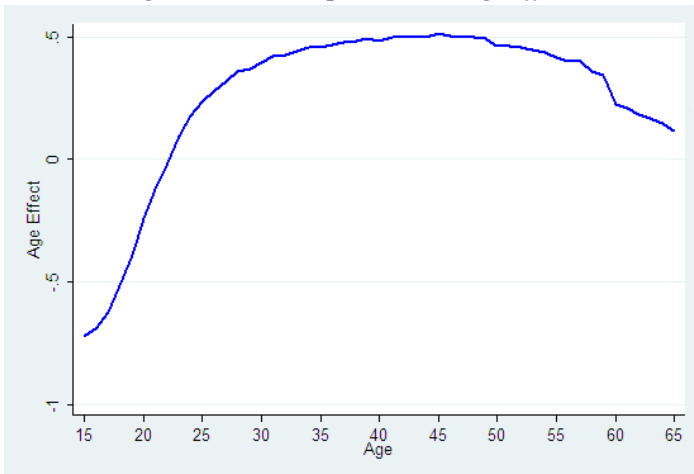


Figure 8.4 Participation rate year effects

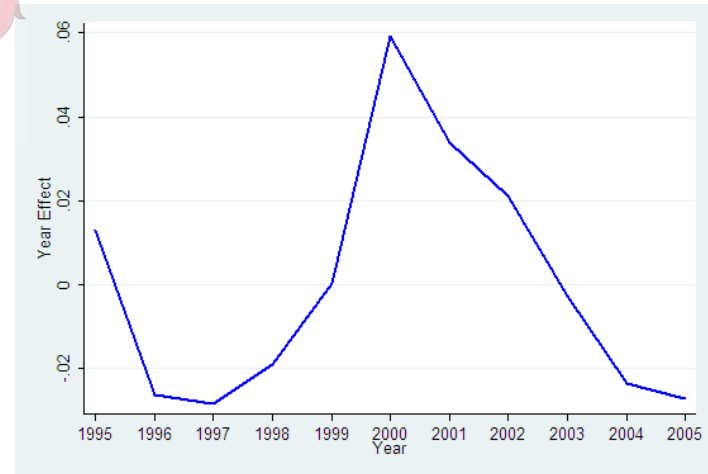


Figure 8.3 Participation rate birth cohort effects



4.2.1 Decomposition of unemployment rate by age, cohort and year

Figure 9 shows the unemployment rate decomposition by birth cohort and population group. Raw cohort data in Figure 9.1 highlights that unemployment is higher for blacks than for whites of all ages and birth cohorts. The figure suggests that youth unemployment is more of a problem, and persists until older ages, for black workers relative to their white counterparts. Nevertheless, white unemployment shows a sudden spike (and a high degree of volatility) at ages younger than 25. The higher unemployment experienced by older labour force participants is a phenomenon that appears to be restricted to the black population.

The year effects for the black and white unemployment decompositions exhibit a large degree of correspondence, except that the white cycle registers a decrease between 1995 and 1996, compared to an increase for the black population.

The unemployment age effects displayed in Figure 9.2 differ markedly between the two population groups. The black age profile is similar to that of the total population (Figure 6.2), except that unemployment shows a small decrease between the ages of 20 and 30 before increasing rapidly after the age of 40. The white age-unemployment profile also demonstrates a decrease in unemployment between the ages of 20 and 30, but the subsequent increase is decidedly more moderate than for the black population. This suggests that youth unemployment is not the focal problem among black cohorts (compared to experiences later in the life cycle), though it is for white cohorts. The latter observation suggests that white cohorts experience transitional problems when they enter the labour market.

Figure 9.3 highlights that black labour market participants from younger generations face higher unemployment rates. This explains why current black youths are often unemployed, despite their relatively favourable position at the beginning of the typical black life cycle (as noted above). Should this poor starting point (as a result of recent entry) deteriorate further with age (as is witnessed in the age profile), the current black generation faces large obstacles. Labour market flows predict deterioration rather than advance as black cohorts age. This trend is consistent with the cohort effects observed for the labour force as a whole.

The situation of white cohorts is strikingly different. The generational profile is considerably lower than that of black cohorts, and stays at a consistent level across most birth years, with a flat gradient. Amongst the very youngest birth cohorts, a sharp increase in white unemployment cohort effects is observed (in contrast to the relatively smooth changes

observed elsewhere): this may be indicative of estimator inaccuracies – possibly caused by the small number of white labour market *participants* sampled from these birth cohorts – or it could signify a new trend emerging among white cohorts. The extent to which each of these factors plays a role will become clear once additional cross sections are added to the pseudo-panel.

It appears that recent white entrants (bar for those very youngest, as referred to above) are not at a great disadvantage (both compared to blacks and previous generations). Yet, the age profile suggests that an initial spell of youth unemployment ensues, which subsequently peters out along the life cycle. The experience of white individuals is thus characterised by “growing pains” followed by a transition into a relatively secure labour market environment.

4.2.2 Decomposition of employment and participation rates by age, cohort and year

Figure 11.1 depicts a fairly comparable participation picture for black and white cohorts. It should be noted, however, that particularly among younger generations, youth participation is much lower among black cohorts than white cohorts. Convergence over the life cycle, however, takes place. This contrasts sharply with the large and persistent differences in employment rates evident in Figure 10.1. Demand for black labour falls short of supply, resulting in a rising surplus in this market segment. This trend is particularly strong amongst young blacks. The previous section showed that this is indeed the case. White labour, while revealing similar supply patterns as their black counterparts, face greater absorption rates, so that labour surpluses are not as prominent in this group.

The decomposition shows that the year-specific employment effects were similar for the black and white population groups, except for the much larger decrease experienced by blacks between 1995 and 1996. The cyclical component of the black and white participation rates also appear to move in unison, although black year effects fluctuate more than those of whites.

The age profiles of participation follow an inverted U-shape, and are very similar for the two races. Participation rates are higher amongst young whites than blacks, but the more rapid inflows for blacks at young ages means that from the age of 30 onwards the age profiles are almost identical, bar for a small persistent gap. Employment-age profiles reveal that white youths experience swift increases in employment probabilities as they move from the age of 18 to 25, after which the employment rate remains high until steadily decreasing after the age

of 50. Black youths face considerably slower initial absorption rates, and the age-employment profile furthermore starts to taper off around an age of 40. This explains the U-shaped form of the black age-unemployment curve, as opposed to the relatively flat profile for whites.

Evidence for early white absorption has been presented elsewhere. Lam & Seekings (2005: 6) show that white Cape Town youths are far more likely to be working than youths from any other population group – this appears to be a “pull” factor, as most of these young white workers are also enrolled in school and come from well-off backgrounds. This contrasts with the experience of other developing countries where youths are “pushed” out of school into employment as a result of household poverty. The slow absorption of young blacks suggests that neither push nor pull factors are traced among this group, despite traditionally higher poverty levels.

Birth cohort profiles for the participation rate indicate that younger generations tend to have a higher proclivity towards labour force participation. Initially black and white profiles move in tandem in a linear movement. For the youngest generations, however, the black population reflects the surge witnessed in the entire labour force (as in the previous section). It would be instructive to consider these effects by gender, given the recent “feminisation of the labour force” (Casale & Posel, 2004:156). Branson’s (2006: 24-26) two-way decompositions reveal that both black males *and* females have contributed to this trend, with recent female entrants in fact exhibiting a more moderate gradient than that of males.

It is furthermore clear that the white group does not drive the recent upward participation trend in the overall population, with this cohort profile tapering off to a horizontal path for the most recent labour market entrants.

The birth cohort profiles of the employment rate for the two population groups, on the other hand, show divergent trends. The white employment rate increases over generations (though a moderate decline is witnessed among the most recent birth cohorts) and black employment decreases consistently for younger cohorts. The increases in white employment therefore moved in the same direction as labour force participation, whereas the stronger increase in black participation rates was met with a *decrease* in employment. This explains why the white unemployment cohort effects remained more or less constant, as opposed to the large increase in the unemployment rate for younger black birth cohorts. The formal economy has clearly not been able to provide jobs for the rapidly expanding labour force, and the burden of this failure has fallen disproportionately on black youths.

The decline of employment birth cohort effects suggests that black unemployment would have increased even in the absence of increasing participation rates. Using the same linearisation as in section 4.1.2, it is evident that 71% of the generational increase in black unemployment, and 56% of the white increase, was driven by an increase in labour force participation, as opposed to a decrease in employment.



Figure 9 Unemployment rate by birth cohort and population group with decompositions, 1995-2005

Figure 9.1 Unemployment rate by birth cohort and age

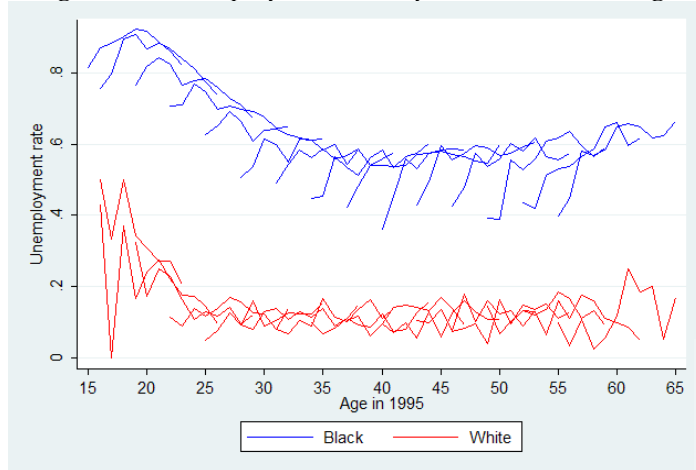


Figure 9.3 Unemployment rate birth cohort effects

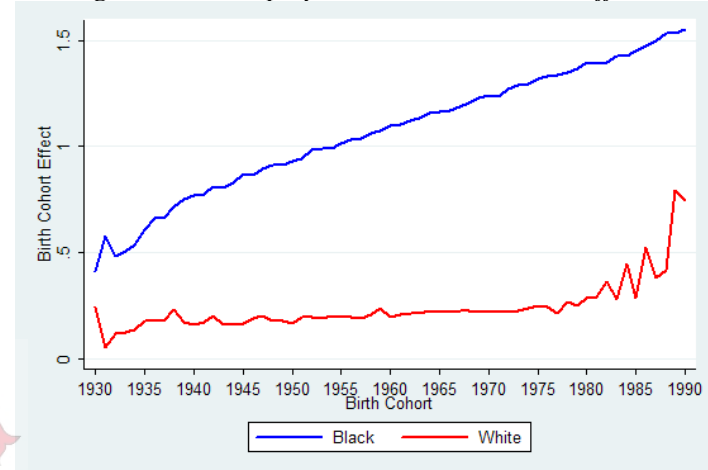


Figure 9.2 Unemployment rate age effects

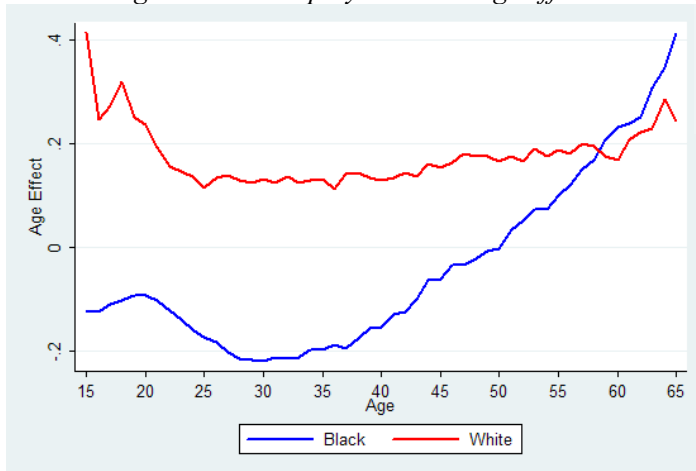


Figure 9.4 Unemployment rate year effects

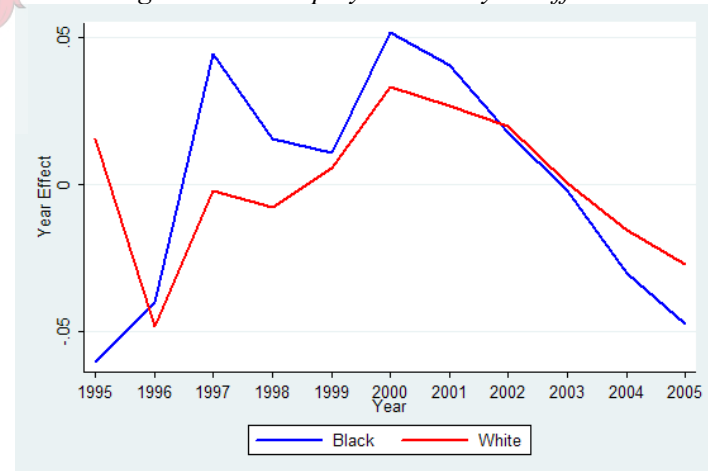


Figure 10 Employment rate by birth cohort and population group with decompositions, 1995-2005

Figure 10.1 Employment rate by birth cohort and age

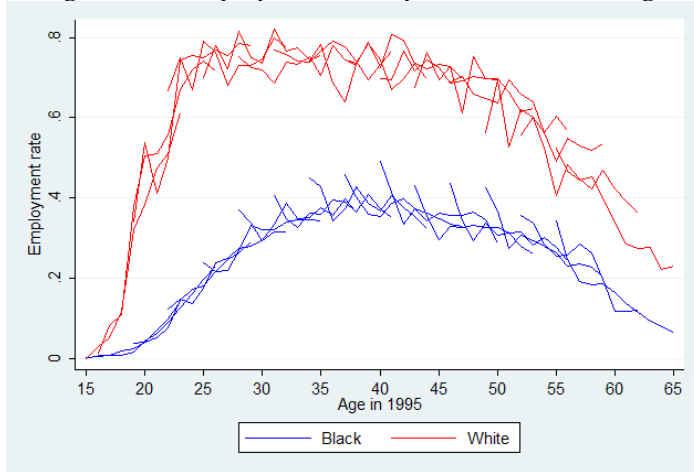


Figure 10.3 Employment rate birth cohort effects

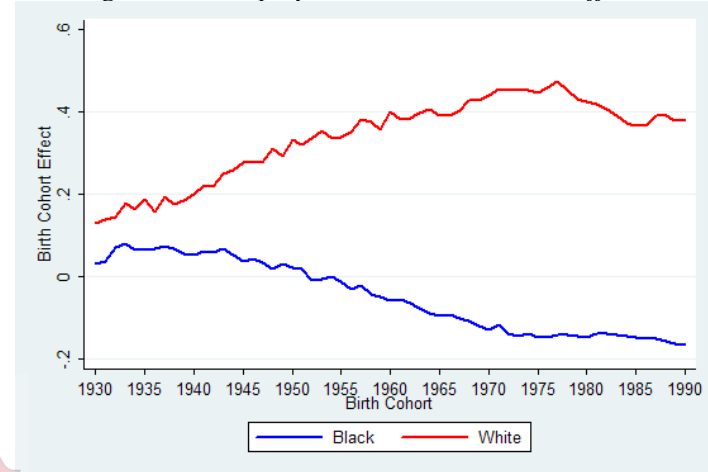


Figure 10.2 Employment rate age effects

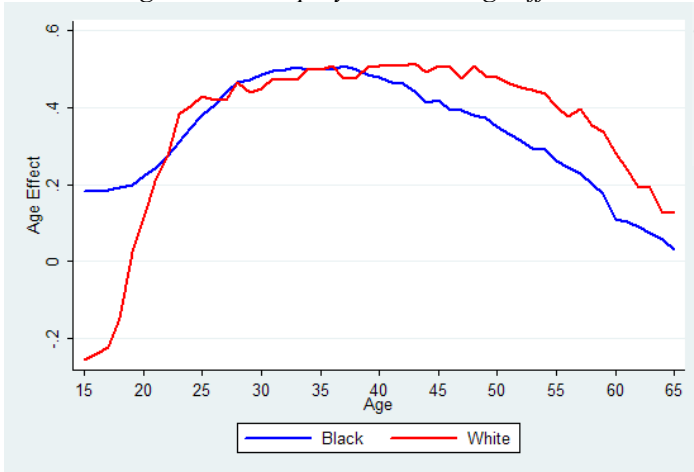


Figure 10.4 Employment rate year effects

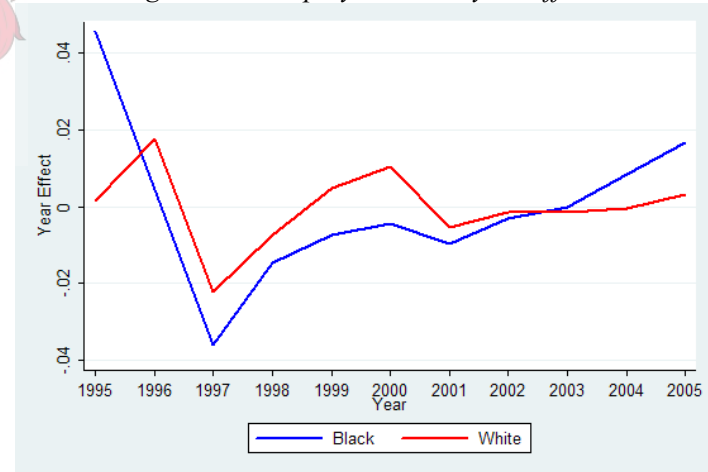


Figure 11 Participation rate by birth cohort and population group with decompositions, 1995-2005

Figure 11.1 Participation rate by birth cohort and age

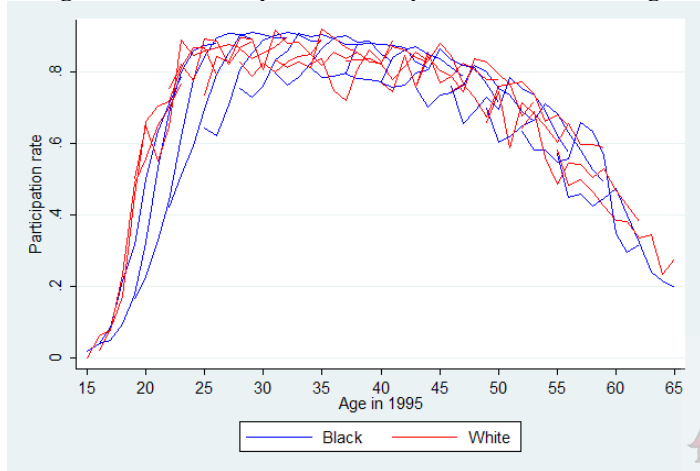


Figure 11.3 Participation rate birth cohort effects

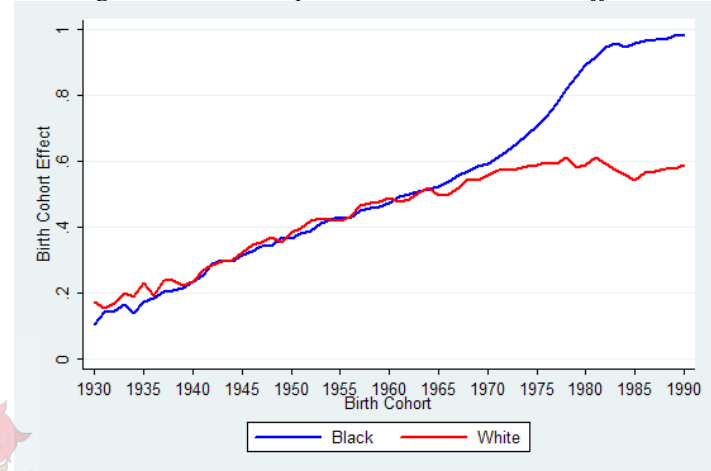


Figure 11.2 Participation rate age effects

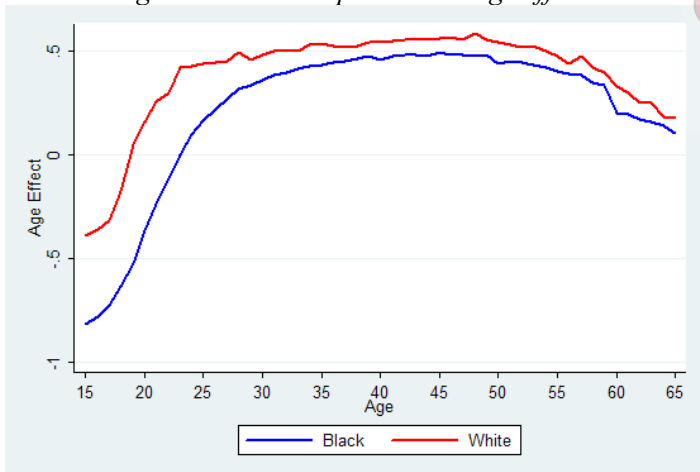
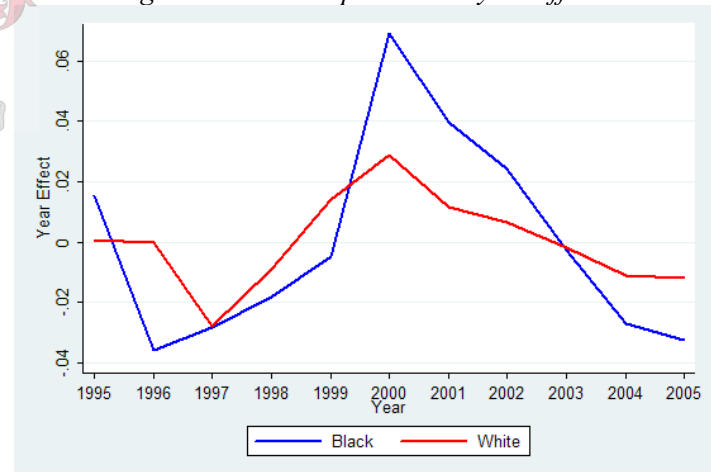


Figure 11.4 Participation rate year effects



4.2.3 Decompositions with controls

The above analysis reveals that the labour market transitions of the black groups most closely represent the movements of the overall working age population. This is expected, given that this is the largest population group within South Africa. It can be observed, however, that the white group has not followed the same trends. Can these population differences, as well as those between generations, be explained by changes in the average level of observable characteristics? By conditioning on variables which exhibit differences across population groups and generations, the “source” of the diverse labour outcomes can be isolated if profiles reveal a more uniform picture for each of these cohorts. The cohort effects of the different sub-populations do not only capture the differential impact of long-term macroeconomic and productivity movements, but also the discriminatory features of the labour market. The latter appear in different forms: differences in educational *quality* (both between population groups and over time) have long been entrenched by separately operated education departments; some cohorts may also have sub-standard educational *attainment* which can be attributed to various circumstances (such as political unrest, poverty and other demographic features). A further source of divergence is taste discrimination, the extent of which can be ascertained most concretely by controlling as exhaustively as possible for observable characteristics. The section which follows undertakes a descriptive analysis to identify potential sources of differential unemployment experiences. Section 4.2.3.2 augments the preceding empirical arguments, by introducing additional controls to the decompositions.

4.2.3.1 Descriptive analysis

Figure 12 highlights differential educational attainment, both across generations and population groups. The sharp changes observed for the most recent birth cohorts should be ignored, as many of these group members are still in transition from one education status to another. What is revealing, however, is the changing face of black education. The oldest generations were unlikely to move beyond primary education, while fewer still progressed to matriculation or tertiary qualifications. For younger cohorts, a sharp decline in primary attainment is accompanied by modest transitions to some high school education, with somewhat larger probabilities of completing secondary schooling and obtaining tertiary education. For all cohorts, most whites have moved beyond primary education. While many members of older cohorts did not complete secondary education, this is rarely the case for their younger counterparts, many of whom also move on to the tertiary level. Human capital theory therefore suggests that the younger generations of both racial groups should be more

productive, and hence face *less* difficulty in becoming employed. This premonition does not come to the fore in the cohort profiles of the preceding analysis. For all generations, whites display a stronger tendency towards matriculation and higher education than blacks. Racial discrepancies, therefore, can potentially be explained by virtue of the favourable educational position of all white cohorts.

Figure 12 Educational attainment, by birth year and population group

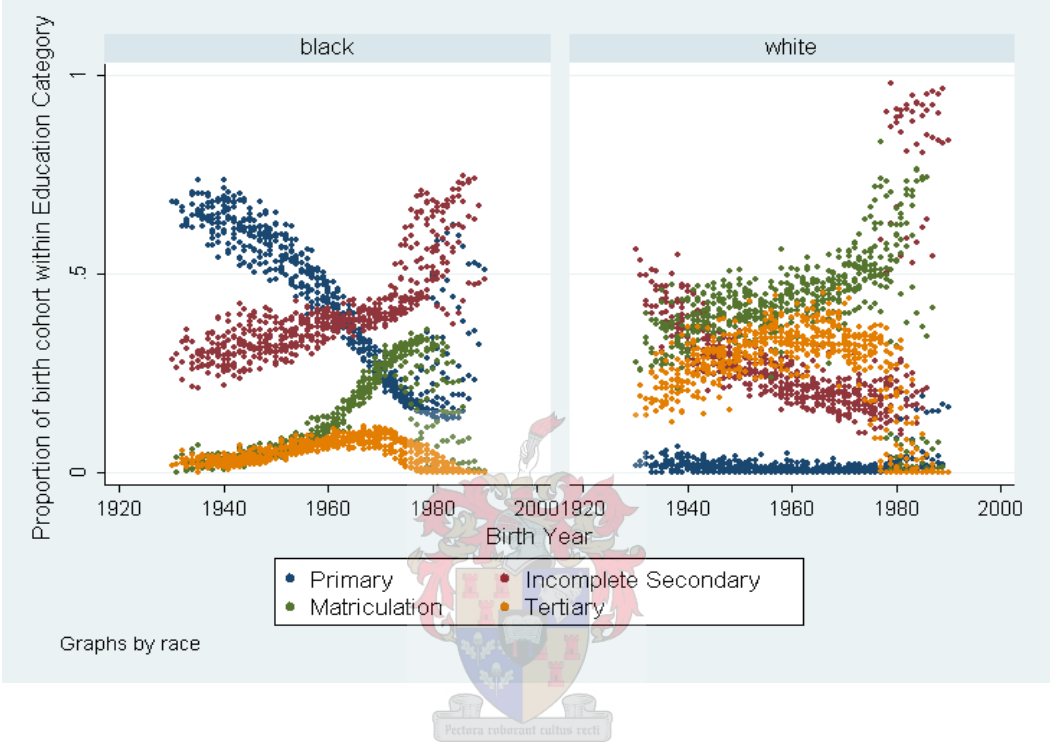


Figure 13 Proportion of over-aged learners, by year and population group

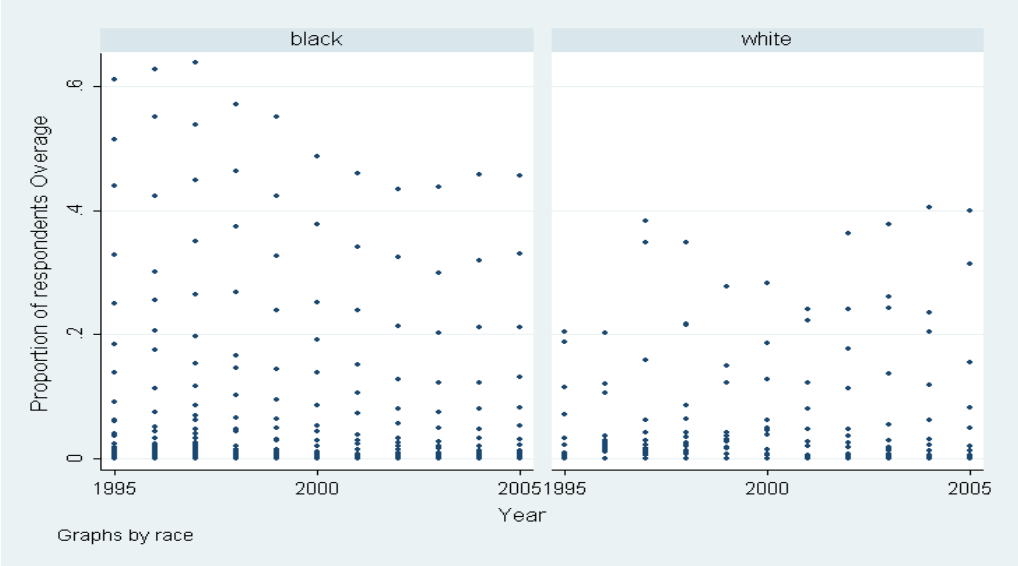


Figure 13 investigates the occurrence of over-aged learners, defined here as individuals beyond the age of 19 presently enrolled in South African schools. Each observation represents a birth cohort in a specific year. It is evident that the large concentration of zero over-aged learners in each year is connected to the fact older cohorts no longer attend school. The highest observation in each year represents the youngest generation. It is clear that black learners are more likely to still be in the schooling system after the age of 19 than white learners, but that the prevalence thereof has been declining. A change in the Department of Education’s policies could therefore have had a substantial effect on the labour market: by streamlining the flow through the schooling system, individuals could have been “forced” into the labour market earlier than would have otherwise been the case, hence increasing the labour force participation rate and possibly also the unemployment rate. From Figure 13 one might expect this to have had a more pronounced effect on black cohorts.

Figure 14 Proportion of individuals living in rural areas, by age and population group

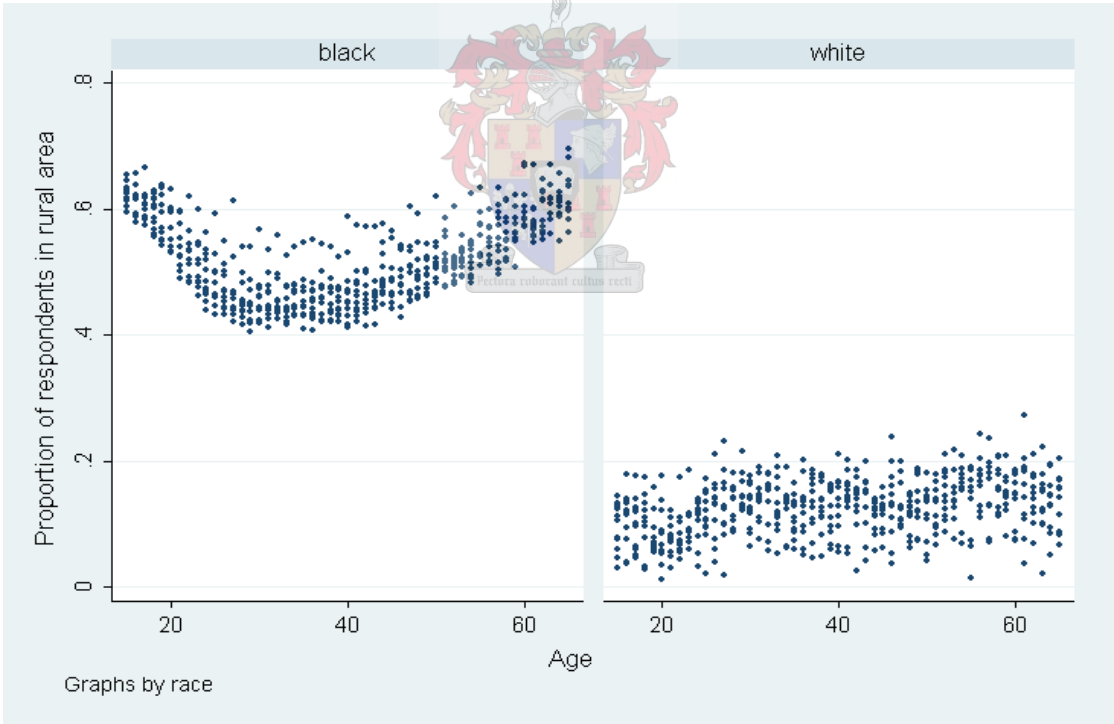


Figure 14 shows the tendency of different age groups to reside (and consequently seek work) in rural areas. High rural unemployment has often been found to be hedged by attachment to households with wage income or which received grants (see for instance Klasen & Woolard, 2005). While residing in a rural area increases the probability of unemployment, coping

mechanisms are varied: they in turn influence household formation, the choice to migrate and also the participation decision. Figure 14 highlights some of these observations. A disproportionate number of younger blacks choose to stay and possibly supply labour in these regions. A definite reduction in this rate for middle-aged individuals highlights the tendency to migrate to cities to find employment. It seems as if many older individuals subsequently return to their rural homes, while the youngest individuals have not yet embarked on the urban route. Posel (2003: 15) hypothesises that insecure labour markets propel this cycle, with migrants not settling permanently in urban areas. As black participants age, their probability of employment declines (see Figure 10.2), and many of these return to rural households, as witnessed in Figure 14. The reasons for delayed migration among the youth are unclear from this picture, though insecure labour markets may be a similar deterrent to urbanisation, as for older individuals. Whites are substantially more urbanised, and uniformly so at all ages. The latter observation explains possible interracial differences; within racial variation in urbanisation (by age) is unlikely to account for much of the increase in white unemployment, as might be expected for black cohorts.

4.2.3.2 Control Variables

Table 3 presents coefficient magnitudes and significance of only the variables used to control for racial and generational differences. A brief scan of Figure 15 to Figure 17 reveals that the inclusion of explanatory variables does not alter the shapes of either the cohort, age or cyclical profiles dramatically, except in isolated cases. Modest gradient adjustments occur, nevertheless. This indicates that differences in geographical location, household composition and even educational attainment does not fully account for the varied racial and generational outcomes.

A discussion of coefficients follows. The presence of non-linear effects of education on labour market outcomes is clearly visible⁹. High shares of tertiary education reduce unemployment rates within both black and white cohorts, yet not significantly. Matriculation and incomplete secondary education in fact increase the probability of unemployment for blacks. This is clarified by investigating the employment and participation models.

For blacks, matriculation provides only a small (and statistically insignificant) increase in employment rates, in comparison to a distinctly large and significant impact on participation. The high black matriculant unemployment rate is therefore fuelled by large increases in

⁹ The discussion considers primary and no education as a reference group

labour supply at this level, combined with smaller gains in employment. This problem is not evident at the tertiary level, where large significant employment gains are registered for blacks, and relatively strong and significant additions to labour supply emerge (though the tertiary coefficient is considerably smaller than that for matriculation). It appears that at this level of education, labour demand and supply is matched. These combine to ease unemployment. This evidence suggests that the oversupply of black matriculants could be reduced by a pass through to tertiary education.

For whites, matriculation has a larger positive impact on employment rates than for blacks, which counters labour surpluses. A similarly important role for tertiary education emerges as for blacks, with employment rates increasing the most at this level. This suggests that this level of attainment is the securest route to finding a job. White participation decisions are very similar at both the matriculation and tertiary levels. These coefficients roughly match those of the employment models, which suggests that the white labour market is closer to equilibrium than the black labour market.

The inclusion of the rural variable is impossible, as it is almost perfectly correlated with the cohort-specific effects (race and birth year in this case), and is consequently dropped due to collinearity concerns. This is of interest in itself, and entails that migration decisions are strongly linked to generational membership and are race specific. To account for geographical variation, cohort provincial shares are thus rather included in the models. These are only significant in any of the white equations, with a strong reduction in unemployment in Gauteng, a high probability of employment as well as participation in the Eastern Cape, and lower participation rates in Mpumalanga. In the first case, white urbanisation is of clear benefit to this group (as Gauteng is a predominantly urban economy). Mpumalanga, in contrast, is a less urbanised province, and white groups respond accordingly with lower levels of activity, or an out-migration of the economically active population.

Variables that explain household composition are significant in almost all cases. White cohorts with high concentrations of household heads experience less unemployment and high participation rates, and are furthermore met with high positive employment prospects for both white and black cohorts. Marriage reduces unemployment for black cohorts, which is supported by higher employment rates. Participation rates are lower in cohorts with high marriage rates, although it may be more instructive to distinguish this effect by gender. A substantial presence of other unemployed household members proves significant for all

cohorts: this scenario exacerbates unemployment vulnerability by way of lower employment probabilities. At the same time, other household members enter the labour market to strengthen the household safety net.

The prevalence of over-aged students has ambiguous effects on unemployment. It however strongly reduces participation, as these individuals remain out of the labour market for longer periods due to grade repetition. This implies that since the introduction of over-age education policies, participation would have increased dramatically. It reduces white employment and increases black employment prospects.

Introducing the lag of the log of wage earnings has the effect of adjusting all the profiles substantially, with some equalisation occurring both across generations and racial groups. It is interesting to note that this variable was insignificant in all regressions (and was therefore excluded from each of the final models), but had the largest ability to absorb cohort effects. Wage outcomes and employment processes are, however, highly intertwined, which calls for some econometric caution. It is wiser to overcome these problems with system instrumental variables estimators (such as 3SLS): a lack of appropriate instruments in this case, however, precludes exploring this relationship in any further depth.

4.2.3.3 Effect of controls on profile differentials

Figure 15 to Figure 17 form the basis of the following discussion. Each panel shows the original decompositions along with the profiles once controls (as above) are introduced. Cohort and age profiles are again adjusted for their respective constants to facilitate interpretation. However, the approach is different for controlled and uncontrolled graphs. Compare the two regressions, the first of which represents the simple decomposition, and the second which introduces the controls:

Equation 14

$$y = c_1 + x_1' \beta_{age1} + x_2' \beta_{cohort1} + x_3' \beta_{year1} + \varepsilon_1$$

$$y = c_2 + z' \beta_{controls} + x_1' \beta_{age2} + x_2' \beta_{cohort2} + x_3' \beta_{year2} + \varepsilon_2$$

The x_i ($i = 1 \dots 3$) are the common dummy variable vectors, while z is the vector of controls introduced subsequently. It is clear that the constants, c_1 and c_2 , might differ substantially, and this suggests a level shift in the profiles. Since the changes of β_i 's effects are of central concern, it is evident that the controlled profiles should not just be scaled by c_2 but by $c_2 + z' \beta_{controls}$. To this end, it is necessary to insert a mean value of z to determine the shift

parameter. A weighted average (according to relative cohort sizes) of the cohorts' productive and demographic characteristics, rather than a simple arithmetic mean, completes the adjustment.

Table 3 Control variable coefficients

	Unemployment		Employment		Participation	
	Black	White	Black	White	Black	White
Incomplete Secondary	0.219 (0.031) **	-0.247 (0.676)	-0.015 (0.779)	0.097 (0.540)	0.110 (0.082) *	0.173 (0.262)
Matric	0.301 (0.047) **	-0.159 (0.773)	0.084 (0.290)	0.265 (0.092) *	0.559 (0.000) ***	0.353 (0.029) **
Tertiary	-0.022 (0.915)	-0.236 (0.681)	0.390 (0.001) ***	0.314 (0.056) *	0.233 (0.065) *	0.352 (0.033) **
EC	0.185 (0.500)	0.007 (0.964)	-0.171 (0.256)	0.191 (0.038) **	-0.199 (0.269)	0.169 (0.054) *
NC	-0.092 (0.886)	-0.140 (0.681)	-0.004 (0.990)	-0.098 (0.621)	-0.379 (0.361)	-0.123 (0.553)
FS	0.051 (0.860)	-0.026 (0.861)	-0.132 (0.410)	0.003 (0.970)	-0.044 (0.825)	0.032 (0.728)
KZN	-0.120 (0.634)	0.021 (0.839)	-0.153 (0.252)	-0.073 (0.324)	-0.217 (0.228)	-0.045 (0.553)
NW	-0.118 (0.689)	-0.072 (0.706)	-0.058 (0.719)	-0.024 (0.819)	-0.048 (0.805)	0.053 (0.641)
GAU	-0.076 (0.767)	-0.238 (0.013) **	-0.001 (0.992)	0.015 (0.781)	-0.088 (0.616)	-0.016 (0.763)
MPU	0.273 (0.363)	-0.279 (0.186)	-0.228 (0.174)	-0.201 (0.111)	-0.204 (0.284)	-0.322 (0.014) **
LIM	0.255 (0.342)	0.062 (0.804)	-0.220 (0.137)	-0.153 (0.350)	-0.077 (0.670)	-0.028 (0.868)
Married	-0.707 (0.000) ***	-0.100 (0.112)	0.171 (0.000) ***	0.106 (0.031) **	-0.263 (0.000) ***	0.083 (0.083) *
Household Head	0.162 (0.097) *	-0.116 (0.030) **	0.166 (0.000) ***	0.255 (0.000) ***	0.435 (0.000) ***	0.232 (0.000) ***
Number of Unemployed in Household	0.250 (0.000) ***	0.483 (0.000) ***	-0.108 (0.000) ***	-0.037 (0.507)	0.102 (0.000) ***	0.289 (0.000) ***
Over-age	-0.047 (0.505)	0.344 (0.054) *	0.115 (0.025) **	-0.777 (0.000) ***	-0.586 (0.000) ***	-0.692 (0.000) ***
N	561	560	561	561	561	561

Robust p-values in parentheses

*significant at 10% **significant at 5% ***significant at 1%

It is evident (in Figure 15.4, Figure 16.4 and Figure 17.4) that in each case year effects are not dramatically altered by controls. This shows that the cyclical variation in labour market outcomes (as well as survey-specific sampling error) is not correlated with educational attainment and demographic factors. This is to be expected, since individual characteristics should not vary in direct relation to the business cycle.

For unemployment, cohort profiles undergo neither any notable shape nor gradient changes (Figure 15.3). Therefore educational attainment does not fully control for generational differences in unemployment (as expected in the descriptive analysis). Age effects are largely unchanged for whites (Figure 15.2). For blacks, the U-shaped profile disappears, with a strong linear upward age trend in unemployment. This may be the result of controlling for educational categories and over-aged learners: the disadvantage which younger cohorts face due to initially low attainment is removed once this conditioning is in place.

Figure 16.3 reveals that controls again do not account for any substantial generational or racial differences in employment, with both population groups undergoing a minor, yet non-uniform downward shift. This might be ascribed to higher educational attainment for younger generations, where the gap is somewhat larger. Once high attainment is removed from contention, recent generational disadvantage is emphasised somewhat. Age profiles are of greater interest. Employment rates of older participants remain unaffected by controls, though youth employment levels are boosted after the introduction of conditional variables. This may again be accounted for by low initial life cycle educational attainment, which, once “removed”, has a smaller restraining effect on youth employment. This observation is consistent with findings for black unemployment.

Figure 17.3 reveals that white generational participation rates are mildly influenced by controls, with an all-round reduction. A more pronounced fall prevails for black profiles, with the largest gap induced for the youngest generations. For this group, the now familiar participation acceleration is subdued, so that this profile resembles that of the uncontrolled white profile more closely. This echoes the premonition that the high educational attainment among particularly younger blacks has fuelled participation. Some racial equalisation therefore occurs, once other characteristics are taken into consideration. Age effects (Figure 17.2) again remain stable for older generations, while an upward participation shift results for the youth. These changes, in combination, suggest a positive role for educational attainment in participation decisions: firstly, recent black labour market entrants’ higher qualifications

possibly drive much of the concurrent increases in labour market activity; secondly, a portion of the low participation rates at young ages may be driven by an initial educational disadvantage relative to older cohorts. All three age profiles underscore this finding. The more dramatic result for participation cohort effects is not mirrored in any of the other profiles. This picture hints towards the educated youth hypothesis, which however requires further verification. The next section embarks on this task.

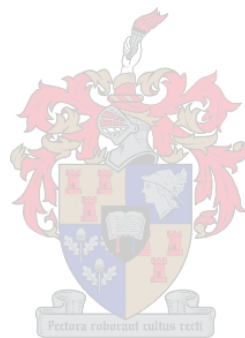


Figure 15 Unemployment rate by cohort with decompositions – with controls, 1995-2005

Figure 15.1 Unemployment rate by birth cohort and age

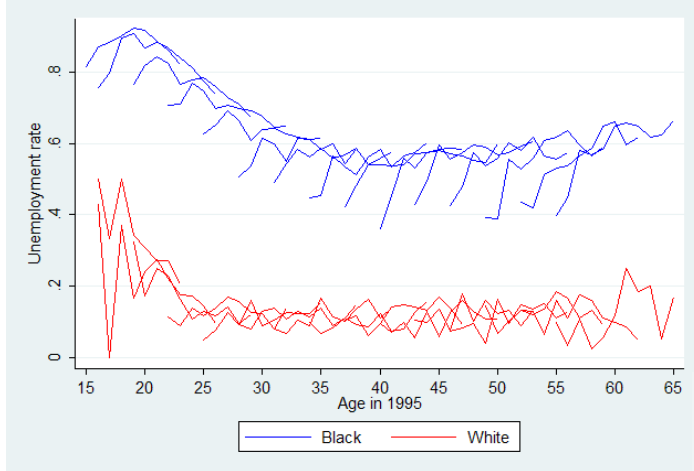


Figure 15.3 Unemployment rate birth cohort effects with controls

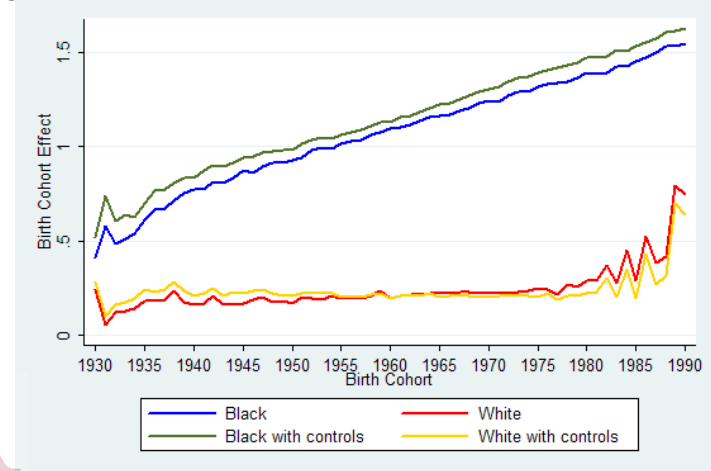


Figure 15.2 Unemployment rate age effects with controls

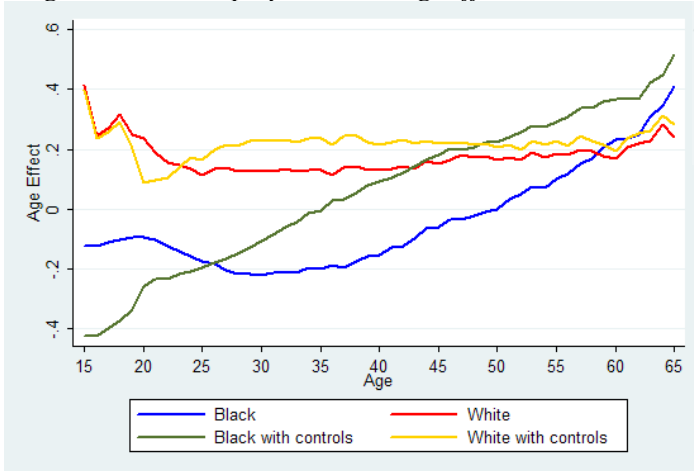


Figure 15.4 Unemployment year effects with controls

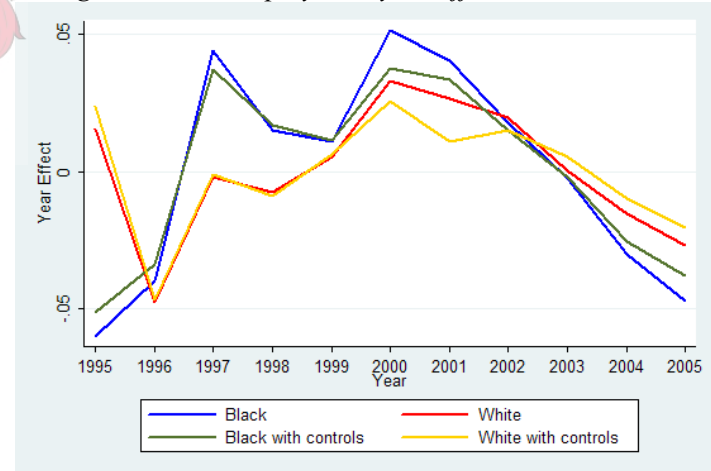


Figure 16 Employment rate by cohort with decompositions – with controls, 1995-2005

Figure 16.1 Employment rate by birth cohort and age

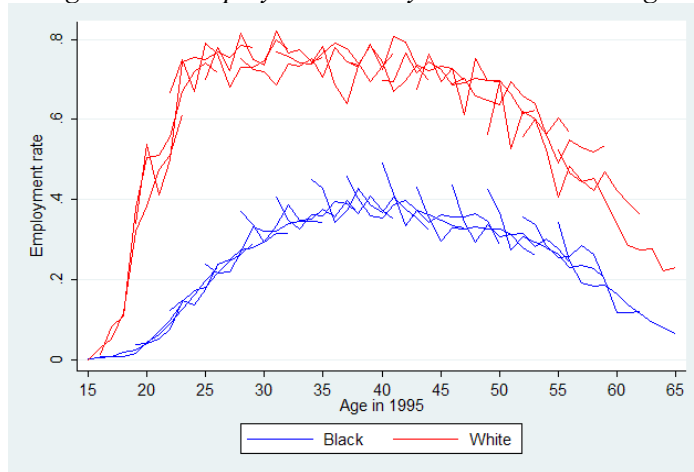


Figure 16.3 Employment rate birth cohort effects with controls

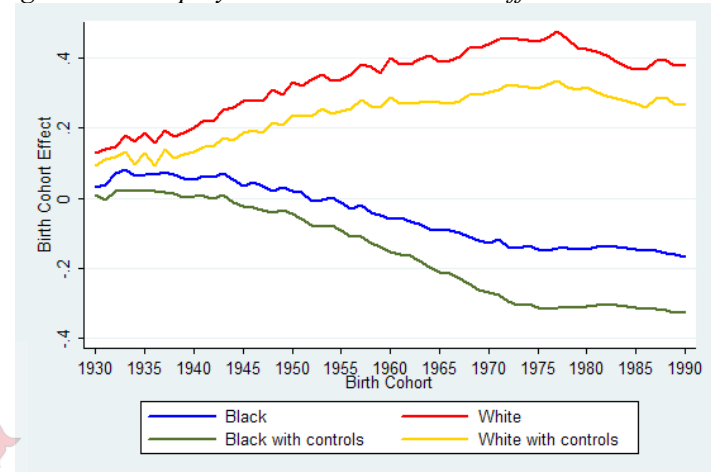


Figure 16.2 Employment rate age effects with controls

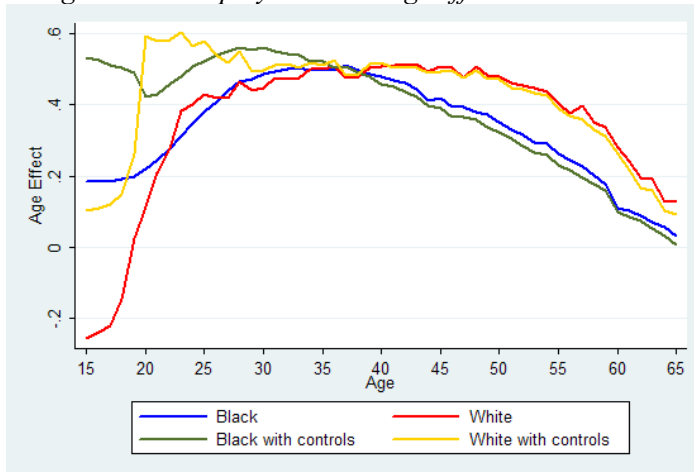


Figure 16.4 Employment rate year effects with controls

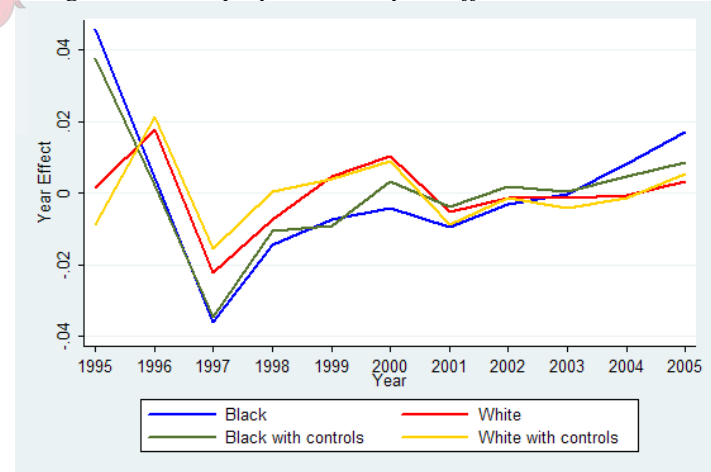


Figure 17 Participation rate by cohort with decompositions – with controls, 1995-2005

Figure 17.1 Participation rate by birth cohort and age

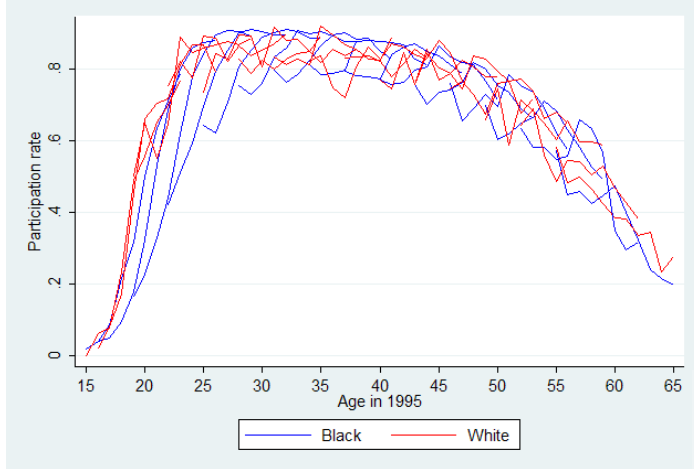


Figure 17.3 Participation rate birth cohort effects with controls

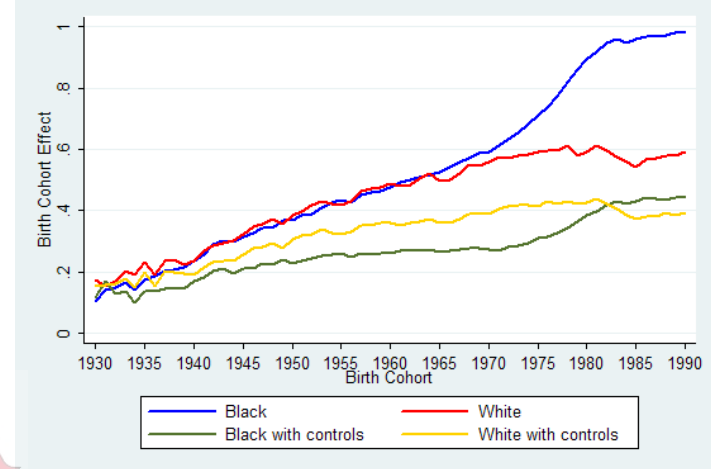


Figure 17.2 Participation rate age effects with controls

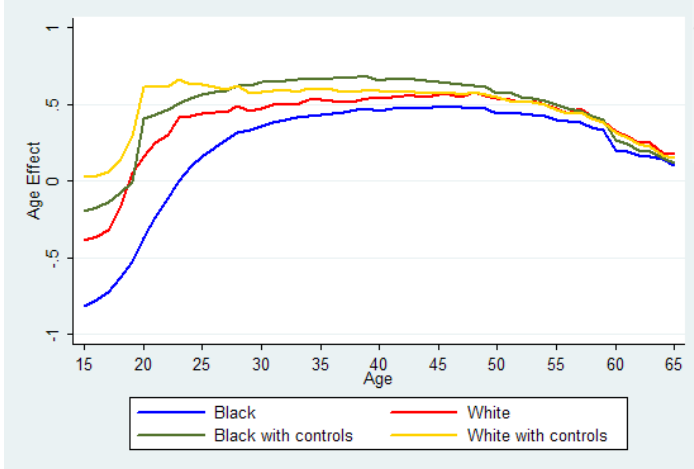
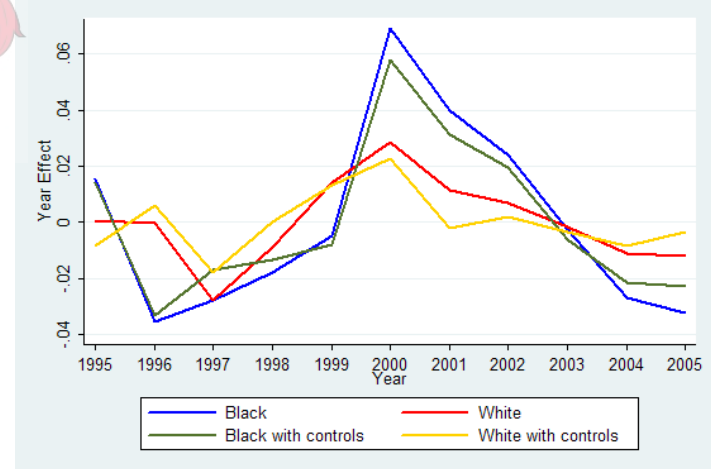


Figure 17.4 Participation rate year effects with controls



4.3 Decompositions by education level

The previous sections highlight that younger generations are at a greater disadvantage compared to their predecessors. This is true, however, against the backdrop of improved educational attainment. The effects of controls in the previous section hinted at some underlying role of education in fuelling participation for certain cohorts. To further uncover these relationships, it is instructive to extend the analysis by considering experiences of educational cohorts.

Branson (2006: 23) shows that, despite the fact that black individuals are entering the labour market earlier, years of education have increased for younger birth cohorts. Therefore faster flows through greater stretches of the education system, rather than exit at low grades (following much repetition), fuel labour supply among the young. This is consistent with the government's intention to eliminate the prevalence of over-aged learners in South African schools (Republic of South Africa, 1995, par 33). It does beg the question, however, whether these accelerated flows have improved the fortunes of individuals *outside* the education system. How has the labour market rewarded different levels of education across generations? Which type of education has born the brunt of rising unemployment? Kingdon & Knight (2005: 35) conclude that between 1995 and 2003 individuals at all education levels experienced higher probabilities of unemployment, yet university graduates were most insulated against this economic trend. Over the same period, Dias & Posel (2006) conclude that education only significantly reduced unemployment after matriculation.

Changing employment prospects for different education levels could reflect one of two phenomena: first, a transformation in the skills structure of labour demand, and secondly changes in the quality of education received by younger generations. The first concern has been documented extensively in the literature. Borhat & Hodge (1999: 368-369) perform sectoral decompositions which show that unskilled employment in South Africa was disadvantaged both by inter-sectoral shifts in economic activity and by intra-sectoral changes in production over the period 1970 to 1995. This phenomenon can be ascribed to structural movements from primary economic activity to a service-orientated economy, but also to increased skills intensity in production even *within* traditional sectors.

The second concern is based largely on the evaluation of a changing education system. South African education has moved from a fragmented, unequally resourced system under Apartheid governments, to a redistributive role following the political transition. Van der

Berg & Burger (2003: 504) highlight that traditionally black schools continue to deliver substantially fewer learners with matriculation endorsements, thereby entrenching poor access to “high return” tertiary education. This is true despite focussed shifts in resource allocation to these schools. They conclude that, in its current state, South African education is unable to address labour market inequalities without an improvement in the management capacity of schools. The latter varies considerably among schools, and consequently perpetuates disparate educational outputs, which feeds through to the labour market.

The cohort effects of black labour market participants for each education category provide an appropriate measure to monitor whether any changes have indeed filtered through to the labour market. It is possible that a longer pseudo-panel is required to establish any clear trends, given the poor progress to date. The unobserved heterogeneity partially absorbs educational quality, but is most successful at capturing an overall picture of the way the same educational attainment has been rewarded in subsequent generations. It is true that a matric obtained by an individual born in 1960 does not possess the same characteristics nor value in the labour market as one attained in 1990. This can be traced back to curriculum differences, but also to supply-demand factors: should many individuals hold the same qualification, the value thereof declines.

Indeed, Wittenberg (2002: 1193-1194) highlights the poor signalling device which a matriculation certificate provided to recruiters in recent times: this can be ascribed to the variable outcomes associated with the same qualification, as well as pronounced levels of this attainment. Anecdotal evidence illustrates this stalemate: on one occasion 39000 individuals applied for 20 low level positions at the University of Cape Town, which crippled the recruitment process (Wittenberg, 2002: 1193-1194). A vicious circle ensues: learners perceive that matriculation secures employment (as may have been true in the past), yet once this qualification is achieved by many, an oversupply of this skill level results in a continued mismatch in the labour market. Powdthavee (2003: 7-8) establishes a strong negative correlation between the completion of either Grades 11 and 12 and happiness levels of South Africans. This leads to the hypothesis that perhaps these learners expected this level of education to improve their income, but that this hope is unwarranted. As a result, misperception leads to low utility levels. These conclusions could explain a high supply of labour at a certain education level, despite a contrasting demand environment.

These misperceptions are also apparent among university graduates. A recent survey of prominent South African firms confirms that new recruits had inflated expectations of the labour market (Paauw et al, 2006: 22). Furthermore, too many students enrolled for “soft” courses (such as the humanities and business degrees), for which labour demand is lower relative to more technical disciplines (Paauw et al, 2006: 16). Graduate unemployment follows, despite the overall insulation referred to above. This burden disproportionately affected black graduates (Dias & Posel, 2006: 4).

The sub-section which follows therefore investigates the effects of differential educational levels on black labour market outcomes. This group is of interest, because a natural experiment of sorts provides a control and treatment group, with cohorts educated in pre- and post-Apartheid schooling systems appearing in the data. One would expect the fortunes of blacks to improve, or at least change to a great extent following the incorporation of former Department of Education and Training Schools into the National Department of Education. The literature reviewed here, however, negates this expectation. Furthermore, the study of this group is motivated by the preceding analyses, which showed that education successfully controlled for some participation spurts. Small cohort sizes preclude comparisons with whites. Within the black group, unemployment decompositions are also at risk of noisy estimates due to small cohort sizes for some educational categories. The improved reliability of estimates for employment and participation rates (bar for some smaller cohorts for more advanced education levels), however, supplement the initial analysis to verify that no improbable results are obtained as a consequence of the grouping strategy. A summary of the cohort sizes used can be found in Table 6. Again, no causal inferences can be made from these analyses, despite the circumstantial evidence and other empirical work cited above, which could serve as potential explanations for the varying outcomes. Cohort and age profiles are again adjusted by the respective constants from the LSDV regressions to aid comparison.

4.3.1 Decomposition of unemployment rate by age, cohort and year

A cursory glance at Figure 18.1 suggests that the distinction between the unemployment rates of different education levels and across birth cohorts is a blurry one. This is probably the product of some measurement error. A closer look, however, reveals that the fortunes of those without education and primary attainment are very similar, with only slight improvements for individuals with some secondary education. It is only at the matriculation level that a departure from the high unemployment rates is evident. This confirms the conclusions of Dias & Posel (2006) that matriculation is the minimum level of education required to counter high

unemployment probabilities. It should be noted, however, that this profile is in no way smooth and shows high volatility both across generations and within cohorts. While measurement error could be blamed for this feature, a prominent point observed in the literature surveyed above would suggest that the highly variable value of matriculation can possibly be ascribed to the dispersed and unpredictable outcomes from different pockets of the schooling system. Finally, tertiary education distinguishes itself as the surest hedge against unemployment, with some older cohorts registering zero unemployment. Figure 18.5 and Figure 18.6 show these relationships more clearly.

Figure 18.2 delivers gradual upward-sloping age-unemployment profiles, with convexity only evident for secondary and tertiary education (though only mildly in the latter case). This suggests that youth unemployment per sé is most serious for individuals with higher levels of education, which is counter to conventional wisdom. The consequent recovery of particularly tertiary education's profile to very flat, low levels, suggests that this is a temporary phenomenon. The graduate unemployment problem may therefore be regarded as a transitory "flow" obstruction; while the movement from tertiary institutions into the job market may be difficult, the rewards and security which follow prove to be the strongest unemployment buffer over the life cycle. Scarring is therefore of little concern for those with tertiary education. These conclusions resemble those found for the *white* population as a whole (Figure 9). In contrast, the early reversal of the matriculation profile at age 40 shows that initial gains are not sustained. After this point, the profile tracks that of lower education levels fairly closely. The dominance of unemployment at all ages for matriculants is somewhat surprising. This question is addressed more comprehensively when unemployment is disentangled into participation and employment components. Profiles for each of the pre-matriculation education levels follow similar trends. They start at relatively low levels of unemployment at a young age, which escalate thereafter. This implies easy entry into positions with a low skills component at a young age, but that exit later in life is also easy. The sharp initial increase in unemployment before age 20 for tertiary education is improbable, and is the product of not only small cohort sizes, but incorrect coding or responses. While it is possible that some individuals would have gained some tertiary training as early as age 15, this is very rare. The spike in the profile can therefore be safely disregarded.

Cohort effects are depicted in Figure 18.3. The order of benefit is as expected, with uneducated cohorts dominating unemployment, and each category progressively experiencing more advantage, with tertiary education proving to be least adversely affected. In each case,

unemployment has risen for younger birth cohorts. Slopes are similar across categories, bar for tertiary education, which displays a more moderate increase. This indicates again that tertiary education is the most effective remedy against unemployment. The noise evident at the endpoints of both the tertiary and matriculation profiles should not be considered as reflective of the true cohort experience, but is evidence of small cohort size.

Cyclical unemployment follows the same direction for each education category (Figure 18.4). It is evident, however, that matriculants, primary education cohorts and the uneducated show a much sharper response to an economic downswing, which manifests in a strong surge in unemployment. Whether this is a survey-specific error or whether one can conclude that poorly skilled workers are worst affected by economic fluctuations, requires further investigation. The uniform turning point evident in the year 2000 is again indicative of unemployment's responsiveness to the business cycle, with a lag of approximately a year.

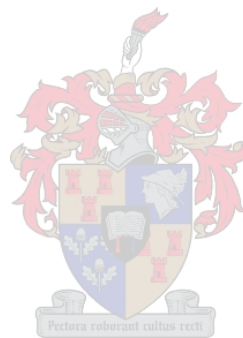


Figure 18 Black unemployment rate by birth and education cohorts, with decompositions, 1995-2005

Figure 18.1 Black unemployment rate by birth cohort, age and education category

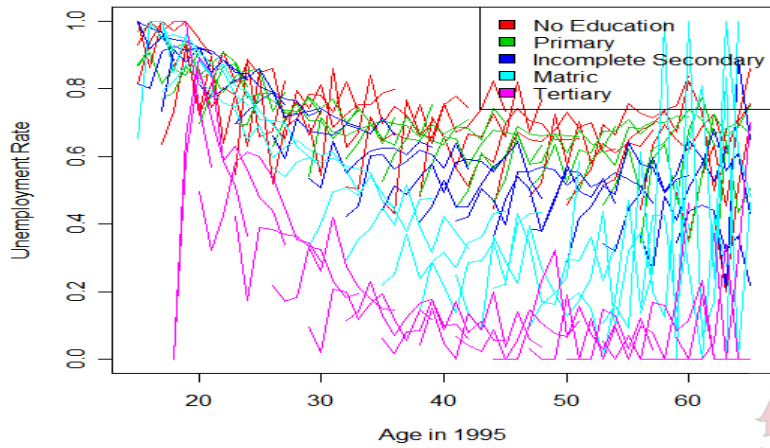


Figure 18.3 Black unemployment rate birth cohort effects, by education category

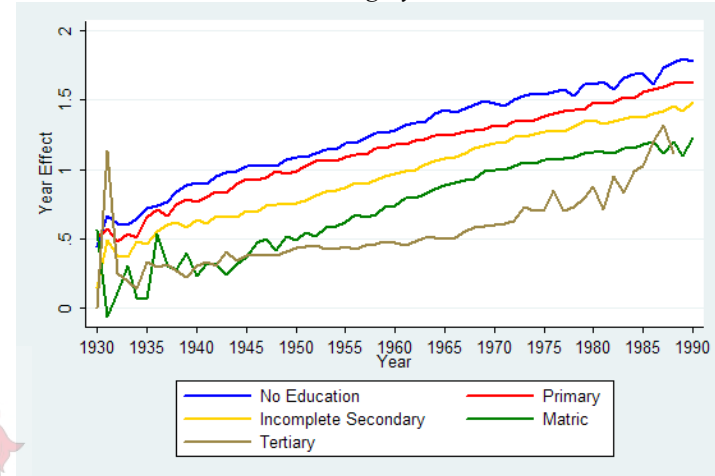


Figure 18.2 Black unemployment rate age effects, by education category

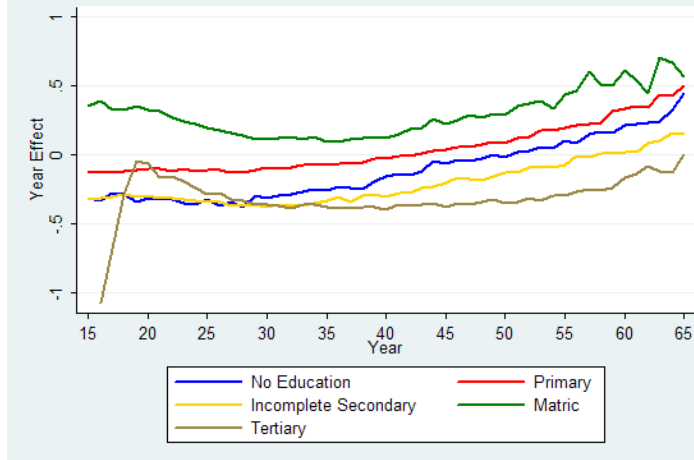


Figure 18.4 Black unemployment rate year effects, by education category

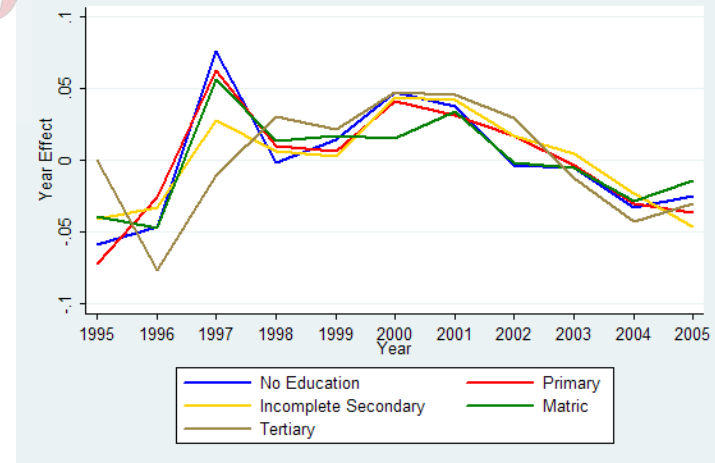


Figure 18.5 Black unemployment rate by birth cohort, age and education category: Primary, Matric and Tertiary Education

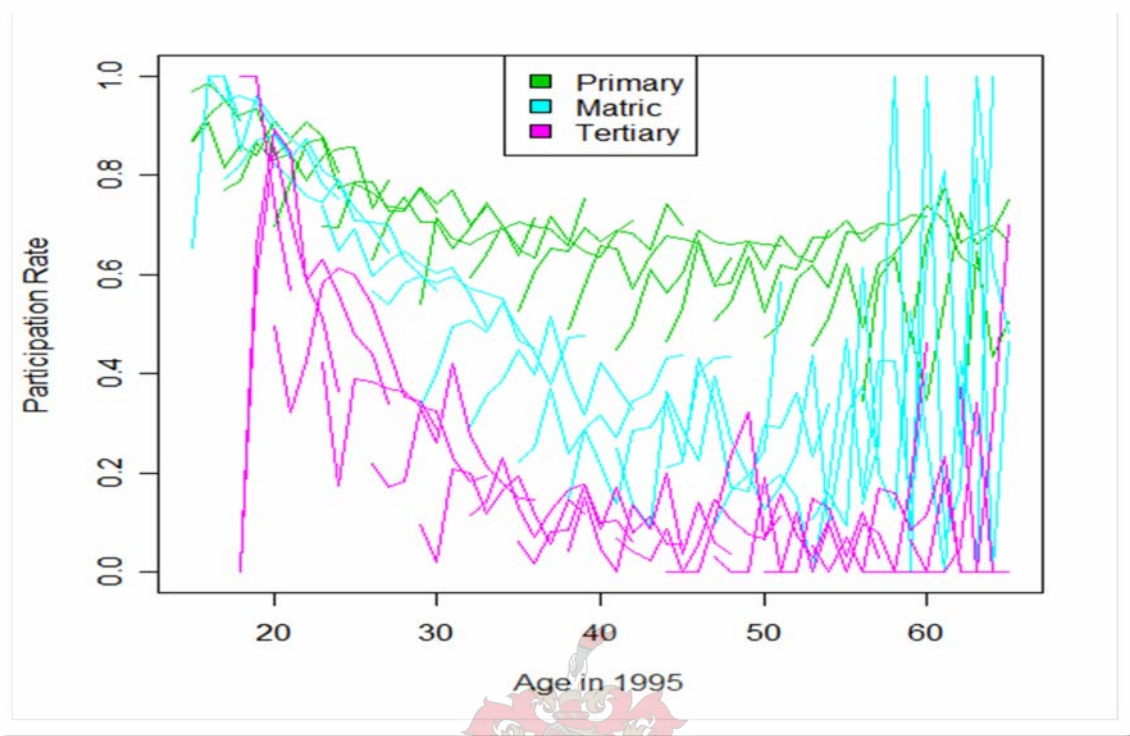
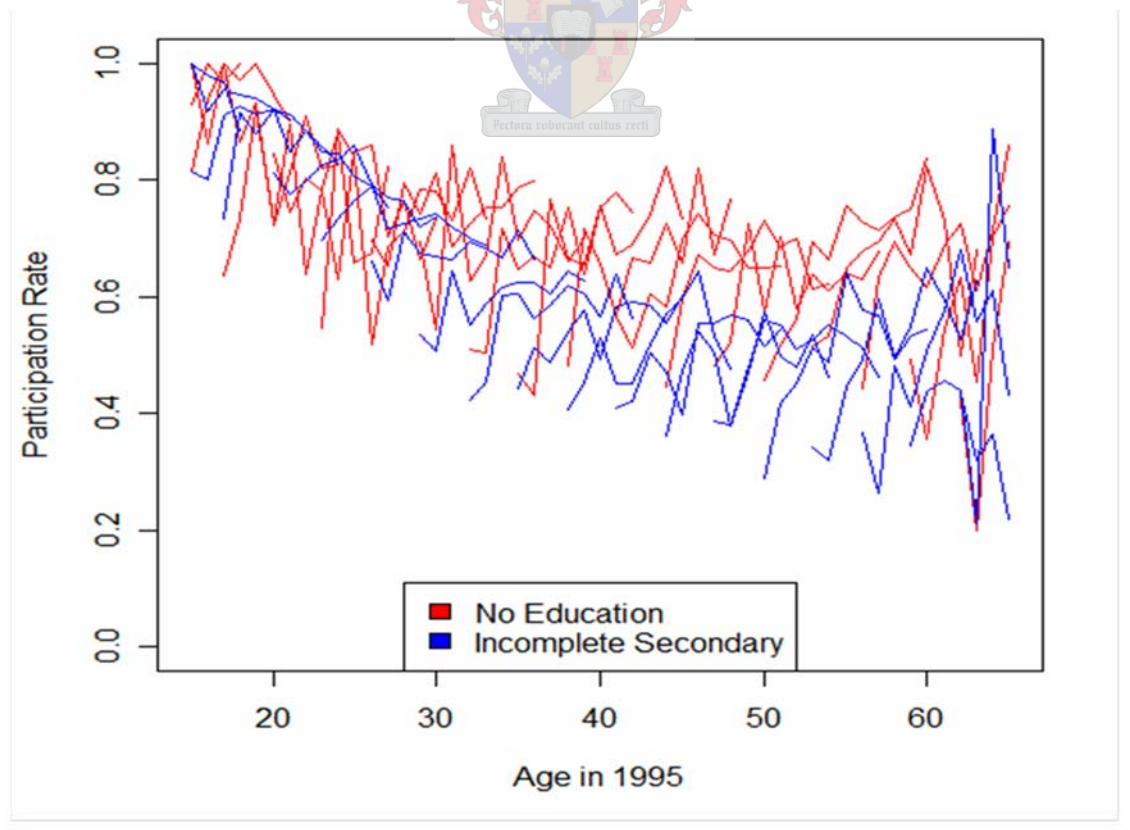


Figure 18.6 Black unemployment rate by birth cohort, age: No Education and Incomplete Secondary Education



4.3.2 Decomposition of employment and participation rates by age, cohort and year

A clearer picture of the scenario in section 4.3.1 emerges once it is broken down into employment and participation components. Figure 19.1 portrays distinct “layers” of employability, with tertiary education cohorts exhibiting the highest absorption rates, followed by matriculants, incomplete secondary, primary and uneducated cohorts. While the three least educated categories move very closely together, it is evident that matriculation offers substantially more opportunities, particularly for older generations. Even further strides are made once tertiary training is received. This separation is not as evident for participation profiles (Figure 20.1), though a somewhat distorted ordering can be discerned in the same vein as for employment rates. It is therefore evident that some education types endow participants with more “employability”, but that labour supply behaviour does not respond proportionally. This explains why less educated individuals face higher unemployment.

Age profiles (Figure 19.2 and Figure 20.2) are revealing. Most noticeable is the very similar life cycle participation behaviour of matriculants and post-matriculants. The difference in these same groups’ employment profiles explains the high levels of life cycle unemployment for matriculants alluded to in the previous section. Matriculants’ labour market flows reflect those of cohorts with higher education: they enter and leave the market at the same rates as graduates. This is clearly a misperception when viewed in the demand context, and signals towards the educated youth hypothesis (though this is not the case for graduates). While few with tertiary education are employed at age 20 (the erratic behaviour prior to that is ignored), strong growth takes place over the next years. The subsequent levels are sustained until just before retirement age. Matriculants exhibit a different pattern, with more employed in earlier years; however a turning point exists prior to age 40, followed by a steady decline. Viewed in relation to the sustained participation profile, it is evident that matriculants suffer large-scale unemployment later in life. This suggests that higher education has lasting impacts on labour market experiences, and that matriculation alone is only beneficial for the first years of a person’s career, after which other skills become relevant. Participation decisions of this group reflect high labour market expectations. This is also mirrored in the high coefficient for matriculation in the participation model above (Table 3). The incomplete secondary education profile is a toned-down version of the matriculation profile. While participation is slightly lower at all ages, employment rates are substantially lower. It is therefore evident that completing secondary school provides more security, but that labour supply

responses do not match the full decline in employment rates for those who do not complete matric. It is interesting to note the late entry of this group, which points to the fact that many of these individuals initially do not have high expectations of their prospects in the labour market. While they may not be classified as unemployed due to inactivity, one might wish to view these younger groups as “hidden unemployed” or disheartened (in the sense of Dinkelmann & Pirouz, 2002). As a consequence, the jobless youth problem may be understated in this case.

Primary and uneducated cohorts enter the labour market very early, presumably in unskilled positions. Push factors presumably predominate at this level. Participation rates increase until just after the age of 20, after which they flatten off to levels slightly below those who have attained some secondary education. The worrying aspect is the early decline in employment rates for this group, to levels well below any of the other categories. This entails a very uncertain middle-age labour market experience.

Cohort effects (Figure 19.3 and Figure 20.3) show a large generational impact on participation, which is relatively uniform across education categories. Tertiary and matric cohorts are supplying more labour at an increasing rate, while incomplete secondary and primary cohorts are doing so at a declining rate. On the demand side, the scenario is not as even: the employability of those with tertiary education has experienced a fairly horizontal shift across generations, with perhaps a moderate decline. The same cannot be said for the other categories, which have experienced a steady decline in employment rates. This drop is particularly sharp for matriculants, which might reflect the fact that more individuals in later generations have joined this group, yet without more vacancies on offer. The fact that this group is entering the market at a growing pace shows that they perceive their prospects in the labour market to be similar to those of tertiary candidates, though this is not the case. This perhaps reflects an over optimism in this group.

Cyclical components (Figure 19.4 and Figure 20.4) are not as clearly defined as was the case for unemployment. What is apparent is that the amplitude of the cyclical participation rate is greatest for the least educated, and most stable for the most educated. This suggests that the uneducated alter their supply of labour most readily, given changing short-term economic conditions. The participation rates, however, seem to move in the opposite direction to the overall business cycles, which is unanticipated. Employment profiles show only very slight cyclical variation (yet in the expected direction).

Figure 19 Black employment rate by birth and education cohorts, with decompositions, 1995-2005

Figure 19.1 Black employment rate by birth cohort, age and education category

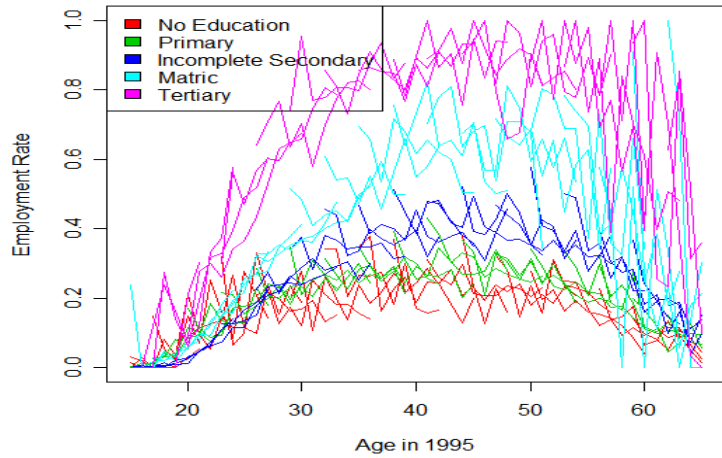


Figure 19.3 Black employment rate birth cohort effects, by education category

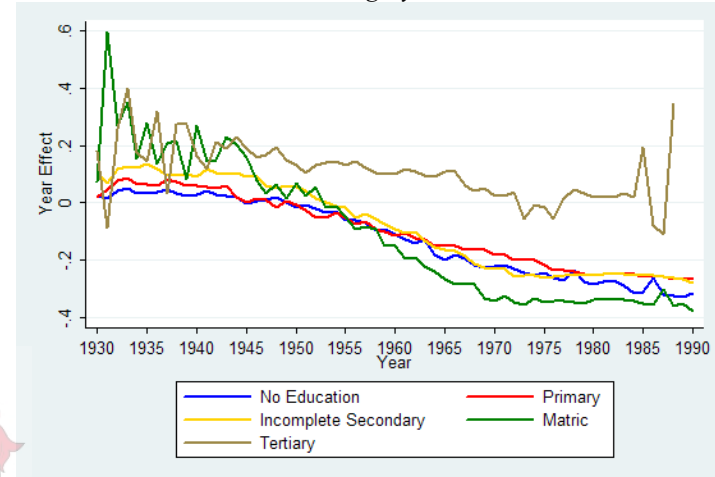


Figure 19.2 Black employment rate age effects, by education category

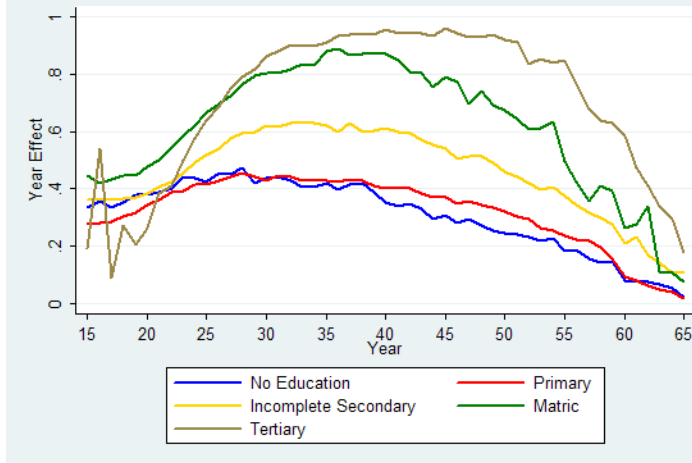


Figure 19.4 Black employment rate year effects, by education category

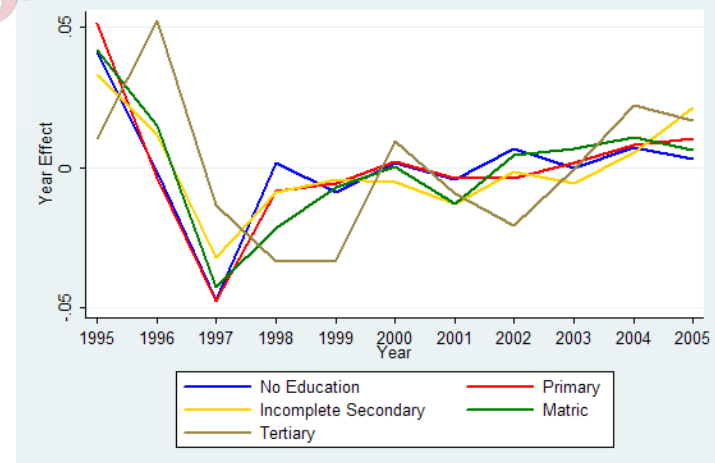


Figure 20 Black participation rate by birth and education cohorts, with decompositions, 1995-2005

Figure 20.1 Black participation rate by birth cohort, age and education category

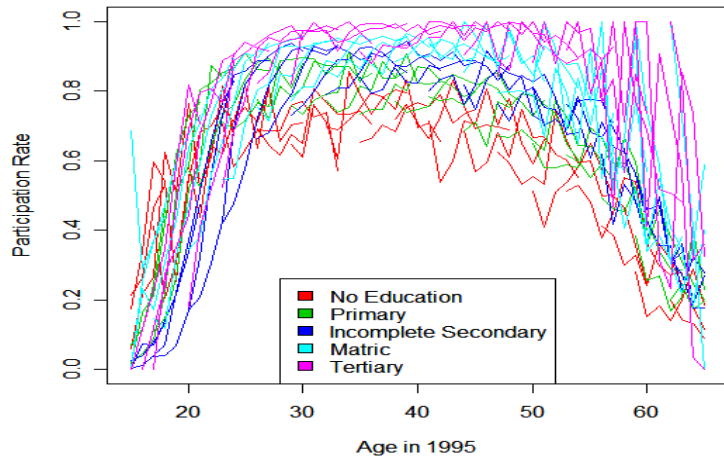


Figure 20.3 Black participation rate birth cohort effects, by education category

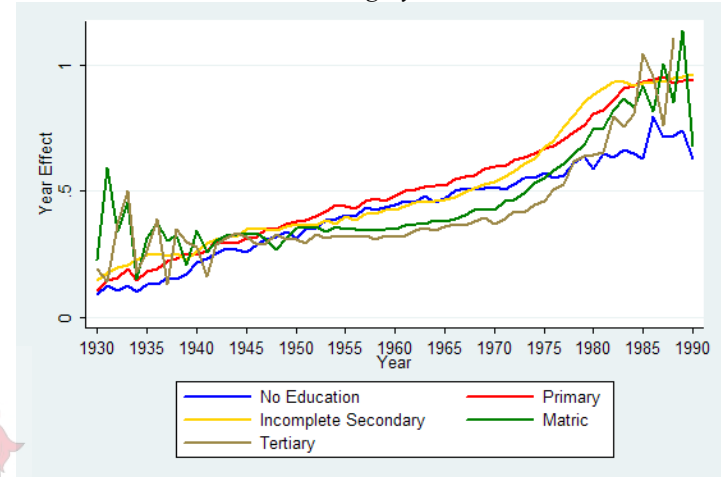


Figure 20.2 Black participation rate age effects, by education category

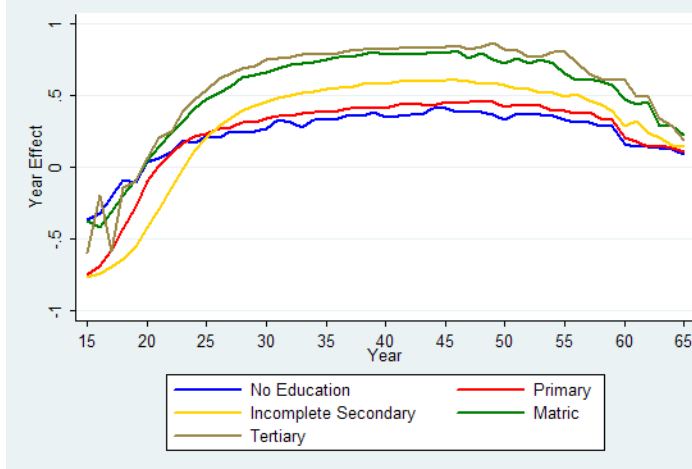
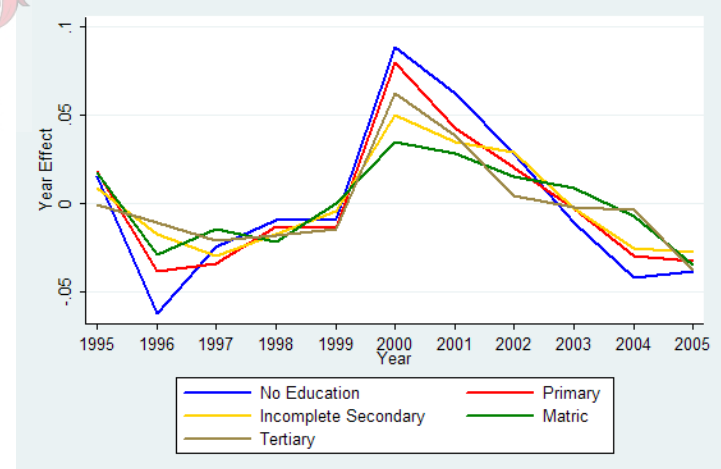


Figure 20.4 Black participation rate year effects, by education category



5 Conclusion

This paper attempted to contribute to the South African unemployment debate by analysing the 17 available post-1994 household survey datasets at the cohort rather than the individual level. Decomposition analyses allowed the identification of a number of important observations, which are highlighted hereafter. Higher unemployment rates faced by the young are predominantly attributable to the disadvantage of entering the labour market more recently, rather than being an age-driven feature. The bulk of this generational disadvantage stems from an increase in the participation rate, as opposed to a decrease in employment opportunities. This participation surge is further clarified by racial comparisons, which reveals that recent black entrants drive much of this trend. Large differences in absorption rates across generations and population groups suggests that whites have had a more secure labour market experience. Youth unemployment in this group quickly subsides, as labour supply matches employment. In contrast, black youths enter the market with generational disadvantages, which are exacerbated by a typical life cycle profile which dictates a further surge in unemployment with age. Much of this can be attributed to life cycle participation patterns which mirror those of whites, coupled with employment flows which leave black participants with fewer job opportunities from their middle age. The potential for labour market scarring is therefore acute in this group. Controls suggest that racial and generational educational differentials have a role to play in this picture.

The black educational pseudo-panel revealed that generational job security and life cycle behaviour of those with tertiary education resembled that of the “secure” white population. It is evident that participants below this attainment threshold more accurately depict the overall black profiles, with optimistic participation patterns and disappointing employment profiles combining to increase recent unemployment. This picture suggests that education had two divergent impacts at two different levels: blacks with tertiary education are able to overcome the disadvantage experienced by the rest of their population group (which is consistent with human capital theory), while increased secondary schooling completion rates have done little more than promote labour market activity without matching gains in employment (which underscores the educated youth hypothesis). The skewness in the returns to education therefore suggests that much needs to be done to overcome the “matric bottleneck”, and to suitably equip labour market entrants.

Throughout, some correspondence between the cyclical variation in unemployment and the official business cycle can be traced from the cross sections. This relationship might, however, be marred by survey-specific sampling error.

The unexplained heterogeneity picked up by pseudo-panels requires further investigation. Can this be viewed as a measure of racial and generational discrimination once productive characteristics are conditioned out? Exploiting a decade of data reveals long-term trends. As additional surveys are appended to the pseudo-panel, it will become possible to distinguish between the truly structural from the temporal features of the economy. Yet, the advent of a reliable pure panel will offer more certain depictions of individual flows through the labour market.

The important South African labour market issues continue to seek suitable means to “share growth”. Past disparities persist, and attempts to equalise these through education have not shown to be favourable upon entry. With difficulties such as the educated youth hypothesis and graduate unemployment in a skills scarce economy (Paauw et al, 2006), it is evident that the “right” solutions may have adverse outcomes. Extensive attainment of matric certificates among younger generations could be hailed a success; yet, one should take cognisance of the realities of the workplace, which requires a set of more developed skills. Furthermore, the educational quality coupled with these higher attainment levels should be taken into account. Supply and demand appears to be balanced for white groups, as well as highly educated blacks; achieving this same equilibrium amongst lower education levels requires a strategic approach to skills development. It may not be possible to offer most of these individuals tertiary qualifications, due to quality constraints at lower education levels. This group nevertheless remains a prominent concern: should labour market scarring affect new entrants, current imbalances could persist well into the future. South Africa’s high unemployment rates affect different communities and generations non-uniformly, and as such policy should be directed at improving the lot of high risk groups. Should the current acceleration in unemployment among these groups become persistent, current challenges could complicate the task of future policymakers.

6 Bibliography

- Arulampalam, W., Booth, A., & Taylor, M. (2000). Unemployment Persistence. *Oxford Economic Papers*, 52, 24-50.
- Baltagi, B. (2005). *Econometric Analysis of Panel Data 3rd Edition*. Chichester: John Wiley & Sons.
- Bhorat, H., & Hodge, J. (1999). Decomposing Shifts in Labour Demand in South Africa. *The South African Journal of Economics*, 67 (3), 348-380.
- Bhorat, H., & Oosthuizen, M. (2005). *The Post-Apartheid South African Labour Market*. DPRU Working Paper 05/93, Development Policy Research Unit, University of Cape Town.
- Branson, N. (2006). *The South African Labour Market 1995-2004: A Cohort Analysis*. SALDRU Working Paper 06/07: Southern Africa Labour and Development Research Unit, University of Cape Town.
- Burger, R., & Woolard, I. (2005). *The State of the Labour Market in South Africa After the First Decade of Democracy*. CSSR Working Paper No 133: Centre for Social Science Research, University of Cape Town.
- Burger, R., & Yu, D. (2006). *Wage trends in post-apartheid South Africa: Constructing an earnings series from household survey data*. Stellenbosch Economics Working Papers 04/06: University of Stellenbosch.
- Cahuc, P., & Zylberberg, A. (2004). *Labour Economics*. Cambridge, MA: The MIT Press.
- Casale, D., Muller, C., & Posel, D. (2004). "Two Million Net New Jobs": A Reconsideration of the Rise in Employment in South Africa, 1995-2003. *The South African Journal of Economics*, 72 (4), 978-1002.
- Casale, D., & Posel, D. (2002). The Continued Feminisation of the Labour Force in South Africa: An Analysis of Recent Data and Trends. *The South African Journal of Economics*, 70 (1), 156-184.
- Clark, A., Georgellis, Y., & Sanfey, P. (2001). Scarring: The Psychological Effect of Past Unemployment. *Economica*, 68, 221-241.
- Clark, A., & Oswald, A. (1996). Satisfaction and comparison income. *Journal of Public Economics*, 61, 359-381.
- Deaton, A. (1985). Panel Data from Time Series of Cross-Sections. *Journal of Econometrics*, 30, 109-126.
- Deaton, A. (1998). *The Analysis of Household Surveys - A Microeconomic Approach to Development Policy*. Baltimore: Johns Hopkins University Press.

- Devey, R., Skinner, C., & Valodia, I. (2006). *Second Best? Trends and Linkages in the Informal Economy in South Africa*. DPRU Working Paper 06/102: Development Policy Research Unit, University of Cape Town.
- Dias, R., & Posel, D. (2006). Unemployment, Education and Skills Constraints in Post-Apartheid South Africa. *DPRU/TIPS Conference on Accelerated and Shared Growth in South Africa: Determinants, Constraints and Opportunities*. Birchwood Hotel: Johannesburg.
- Dinkelmann, T., & Pirouz, F. (2002). Individual, Household and Regional Determinants of Labour Force Attachment in South Africa: Evidence from the 1997 October Household Survey. *The South African Journal of Economics*, 70 (5), 865-891.
- Graham, C., & Pettinato, S. (2002). Frustrated Achievers: Winners, Losers and Subjective Well-Being in New Market Economies. *The Journal of Development Studies*, 38 (4), 100-140.
- Gregg, P., & Tominey, E. (2005). The wage scar from male youth unemployment. *Labour Economics*, 12, 487-509.
- Grün, C. (2004). *Racial and Gender Wage Differentials in South Africa: What can Cohort Data tell?* Retrieved from http://www.wits.ac.za/sebs/downloads/cohort_data.pdf.
- Hsiao, C. (2003). *Analysis of Panel Data - Second Edition*. Cambridge: The Press Syndicate of the University of Cambridge.
- Inoue, A. (2005). *Efficient Estimation and Inference in Linear Pseudo-Panel Data Models, May 2005*. Retrieved from Texas A&M Department of Economics Spring Workshop: <http://econweb.tamu.edu/workshops/Econometrics/Atsushi%20Inoue.pdf>.
- Keswell, M., & Poswell, L. (2004). Returns to Education in South Africa: A Retrospective Sensitivity Analysis. *The South African Journal of Economics*, 72 (4), 834-860.
- Kingdon, G., & Knight, J. (2004). Unemployment in South Africa: The Nature of the Beast. *World Development*, 32 (3), 391-408.
- Kingdon, G., & Knight, J. (2005). *Unemployment in South Africa, 1995-2003: causes problems and policies*. Retrieved from Global Poverty Research Group Working Paper 010: <http://www.gprg.org/pubs/workingpapers/pdfs/gprg-wps-010.pdf>.
- Kingdon, G., & Knight, J. (2006). The measurement of unemployment when unemployment is high. *Labor Economics*, 13 (3), 291-315.
- Klasen, S., & Woolard, I. (2005). Surviving Unemployment without State Support: Unemployment and Household Formation in South Africa. *CSAE Conference on "Reducing Poverty and Inequality: How can Africa be included?"*. St Catherines College, Oxford: Centre for the Study of African Economics.
- Lam, D., & Anderson, K. (2001). Women's Schooling, Fertility and Investments in Children in South Africa. *"Welfare, Demography and Development" Conference*. University of Cambridge, 11-12 September 2001.
- Lam, D., & Seekings, J. (2005). Transitions to Adulthood in Urban South Africa: Evidence from a Panel Survey. *IUSSP General Conference*. Tours, France.

- Leibbrandt, M., & Mlatsheni, C. (2004). Youth in Sub-Saharan Labour Markets. *DPRU/TIPS Conference on African Development and Poverty Reduction: The Macro-Micro Linkage*. Somerset West.
- Ljungqvist, L. (2000). Economic Underdevelopment: The Case of a Missing Market for Human Capital. In P. Bardhan, & C. Udry (Eds.), *Readings in Development Economics, Vol 1: Micro-Theory* (pp. 123-144). Cambridge, MA: The MIT Press.
- Lohr, S. (1999). *Sampling: Design and Analysis*. Pacific Grove: Brooks/Cole Publishing Company.
- Louw, M., van der Berg, S., & Yu, D. (2006). *Educational attainment and intergenerational social mobility in South Africa*. Stellenbosch Economic Working Papers 09/06: University of Stellenbosch.
- Mlatsheni, C., & Rospabe, S. (2002). *Why is Youth Unemployment so High and Unequally spread in South Africa?* Development Policy Research Unit Working Paper 02/65: DPRU, University of Cape Town.
- Oosthuizen, M., & Borat, H. (2005). *The post-apartheid South African Labour Market*. DPRU Working Paper 05/093: Development Policy Research Unit, University of Cape Town.
- Paauw, K., Oosthuizen, M., & van der Westhuizen, C. (2006). Graduate Unemployment in the Face of Skills Shortages: A Labour Market Paradox. *DPRU/TIPS Conference: ASGISA: Determinants, Constraints and Opportunities*. Birchwood Hotel, Johannesburg 18-20 October 2006.
- Posel, D. (2003). Have Migration Patterns in post-Apartheid South Africa Changed? *Conference on African Migration in Comparative Perspective*. Johannesburg (4-7 June).
- Powdthavee, N. (2003). *Are there Regional Variations in the Psychological Costs of Unemployment in South Africa?* Retrieved from Warwick University: <http://129.3.20.41/eps/lab/papers/0310/0310006.pdf>.
- R Development Core Team. (2005). *R: A language and environment for statistical computing* ISBN 3-900051-07-0, Version 2.2.0. Retrieved from R Foundation for Statistical Computing, Vienna, Austria: <http://www.R-project.org>.
- Raum, O., & Roed, K. (2002). *Do Business Cycle Conditions at the Time of Labour Market Entry Affect Future Unemployment?* Memorandum No 12/2002: Department of Economics, University of Oslo.
- Republic of South Africa. (1995). *Meeting the Commitment to Free and Compulsory General Education*. Retrieved from Chapter 13, White Paper on Education and Training, Notice 196 of 1995, Department of Education: http://www.polity.org.za/html/govdocs/white_papers/educ5.html.
- South African Reserve Bank (SARB). (2006). *Quarterly Bulletin - June 2006 No 240*. Pretoria: South African Reserve Bank.
- STATA. (2006). *FAQ: Interpreting the intercept in the fixed-effects model*. Retrieved from <http://www.stata.com/support/faqs/stat/xtreg2.html>.

- Urdal, H. (2006). A Clash of Generations? Youth Bulges and Political Violence. *International Studies Quarterly*, 50, 607-629.
- Van der Berg, S., & Burger, R. (2003). Education and Socio-Economic Differentials: A Study of School Performance in the Western Cape. *The South African Journal of Economics*, 71 (3), 496-522.
- Verbeek, M., & Nijman, T. (1992). Can Cohort Data be Treated as Genuine Panel Data? *Empirical Economics*, 17, 9-23.
- Wittenberg, M. (2002). Job Search in South Africa: a Non-Parametric Analysis. *The South African Journal of Economics*, 70 (8), 1163-1197.
- Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.



Appendix 1: Code to compare GMM and FE Estimators

The following code was written in R version 2.2.0 (see R Development Core Team, 2005) to compare the Fixed Effects Estimator to the GMM estimator of Inoue (2005).

```
function (data,timevar,groupvar,ifvar,crit,size,depvar)
{
#data must be aggregated by cohort and not in its raw individual-level format
#data must be sorted first by birth cohort then by time period, "index" below should perform this order
#data is a data matrix or dataframe, which will be referred to and modified
#timevar is the time index in "data", usually "year"
#groupvar is the cohort index in "data", usually "birth year"
#ifvar (also in "data") is used to isolate a certain subpopulation, for instance "race"
#crit is the value to evaluate the ifvar at, for instance "white" or "black". NOTE: this must within " "
#size is a vector, which contains each of the cohort sizes - choose the variable in "data"
#depvar is the dependent variable in "data"
#Regressors cannot be specified for this programme: they remain the same for each decomposition, and are specified
within the programme

#ensure that the original dataset is maintained
data2<-data
attach(data2)

#Create an index of the "complete" panel
Sind<-min(groupvar):max(groupvar)
Tind<-min(timevar):max(timevar)
STind<-(rep(Sind,length(Tind))[order(rep(Sind,length(Tind)))]*10000 + (rep(Tind,length(Sind))))

#limit dataset to subpopulation
data3<-data2[ifvar==crit,]
detach(data2)
attach(data3)
#create an index of the data2 actually still in the data2set and sort the dataset to match the panel structure
index<-(groupvar[ifvar==crit]*10000)+ timevar[ifvar==crit]
```

```

data3<-cbind(data3,index)
detach(data3)
rm(index)
attach(data3)
data4<-data3[sort(order(index)),]
detach(data3)
attach(data4)

#Prepare variables in model, according to criteria
depvar2<-as.vector(depvar[ifvar==crit])
depvar2<-depvar2[sort(order(index))]
#Base group: Age 65 (m_age51) and birth year 1930 (cohort1); no constant is specified
W<-cbind(m.age1,m.age2, m.age3, m.age4, m.age5, m.age6, m.age7, m.age8, m.age9, m.age10, m.age11, m.age12, m.age13,
m.age14, m.age15, m.age16, m.age17, m.age18, m.age19, m.age20, m.age21, m.age22, m.age23, m.age24, m.age25, m.age26,
m.age27, m.age28, m.age29, m.age30, m.age31, m.age32, m.age33, m.age34, m.age35, m.age36, m.age37, m.age38, m.age39,
m.age40, m.age41, m.age42, m.age43, m.age44, m.age45, m.age46, m.age47, m.age48, m.age49, m.age50,yrstar3, yrstar4,
yrstar5, yrstar6, yrstar7, yrstar8, yrstar9, yrstar10, yrstar11)
W<-W[sort(order(index)),]

#Create Demeaning Matrices
T<-length(Tind)
S<-length(Sind)
lvecT<-rep(1,T)
lvecS<-rep(1,S)
M1<-diag(1,S)
M2<-diag(1,T) - (1/T)*lvecT%*%t(lvecT)
M<-kronecker(M1,M2)

#Reduce demeaning Matrices, so that they correspond with unbalanced panels
y<-rep(0,length(STind))
for(z in 1:length(STind))
{
y<-y+replace(y,z,sum(index==STind[z]))
}
clear1<-y==0
clear2<-1:length(STind)
clear<-clear1*clear2

M<-M[-clear,-clear]

```




```

#Create Weighting Matrix
Size<-as.matrix(size[ifvar==crit])
Size<-Size[sort(order(index))]
N<-rep(sum(Size),length(Size))
pivec<- Size/N
pivec<-1/pivec
PIinv<-diag(as.vector(pivec))

#Obtain GMM Solution
library(MASS)
A<-t(M)%*%PIinv%*%M
MAM<-t(M)%*%ginv(A)%*%M
GMM<-solve(t(W)%*%MAM%*%as.matrix(W))%*%t(W)%*%MAM%*%as.matrix(depvar2)

#Obtain Fixed effects solution (Wooldridge, 2002: 269)
FE<-solve(t(W)%*%t(M)%*%M%*%W)%*%t(W)%*%t(M)%*%M%*%depvar2

#Create Variance Covariance Matrix of GMM Estimator (for computational ease, the estimator of sigma, is sigmaFE, as
shown in Inoue (2005: 8)). This only affects T-statistics, and not the F-statistic proposed in the text above
uvec<-depvar2 - W%*%FE
sigma2<-(1/ (S*(T-1)-dim(W)[2]))*t(uvec)%*%M%*%uvec
VGMM1<-solve(t(W)%*%t(M)%*%ginv(t(M)%*%PIinv%*%M)%*%M%*%W)
VGMM2<- matrix(sigma2/N[1],nrow=nrow(VGMM1),ncol=ncol(VGMM1))
VGMM<- VGMM2*VGMM1
seGMM<-sqrt(diag(VGMM))
tGMM<-GMM/seGMM
pGMM<-dt(tGMM,S*(T-1)-dim(W)[2])

#Create Variance Covariance Matrix of FE Estimator - this is Inoue (2005: 8)'s version, bar for the estimate of
sigma, which is replaced by sigmaFE, and the additional M's to account for the incomplete panel
VFE1rob<-solve(t(W)%*%t(M)%*%M%*%W)%*%t(W)%*%t(M)%*%PIinv%*%M%*%W%*%solve(t(W)%*%t(M)%*%M%*%W)
VFE2rob<- matrix(sigma2/N[1],nrow=nrow(VFE1rob),ncol=ncol(VFE1rob))
VFErob<-VFE1rob*VFE2rob
VFE1<-solve(t(W)%*%t(M)%*%M%*%W)
VFE2<- matrix(sigma2,nrow=nrow(VFE1),ncol=ncol(VFE1))
VFE<-VFE1*VFE2
seFE<-sqrt(diag(VFErob))
#Note that T-statistics are based on the robust fixed effects variance estimator proposed by Inoue (2005)

```

```

tFE<-FE/seFE
tFE<-abs(tFE)
pFE<-pt(tFE,df=S*(T-1)-dim(W)[2],lower.tail=FALSE)

#Recover unmodelled year effects
GMMy95<-(9*GMM[59])+(8*GMM[58])+(7*GMM[57])+(6*GMM[56])+(5*GMM[55])+(4*GMM[54])+(3*GMM[53])+(2*GMM[52])+GMM[51]
GMMy96<-
-((10*GMM[59])+(9*GMM[58])+(8*GMM[57])+(7*GMM[56])+(6*GMM[55])+(5*GMM[54])+(4*GMM[53])+(3*GMM[52])+(2*GMM[51]))

FEy95<-(9*FE[59])+(8*FE[58])+(7*FE[57])+(6*FE[56])+(5*FE[55])+(4*FE[54])+(3*FE[53])+(2*FE[52])+FE[51]
FEy96<-
-((10*FE[59])+(9*FE[58])+(8*FE[57])+(7*FE[56])+(6*FE[55])+(5*FE[54])+(4*FE[53])+(3*FE[52])+(2*FE[51]))

yearGMM<-c(GMMy95,GMMy96,GMM[51:59])
yearFE<-c(FEy95,FEy96,FE[51:59])

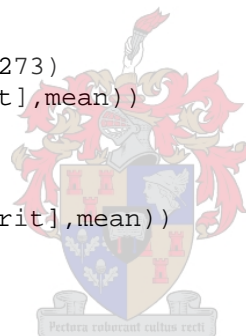
#Recover Cohort Effects, as per Wooldridge (2002: 273)
depmeans<-as.vector(by(depvar2,groupvar[ifvar==crit],mean))
regmeans<-matrix(0,nrow=length(depmeans),ncol(W))
for(c in 1:ncol(W))
{
regmeans[,c]<-as.vector(by(W[,c],groupvar[ifvar==crit],mean))
}
cohortGMM<-depmeans - regmeans%%GMM
cohortFE<-depmeans - regmeans%%FE

#Create Test Statistics for equality of coefficients
F1.stat<-((FE/GMM)^2)*(diag(VGMM1)/diag(VFE1rob))
pF1<-pf(F1.stat,1,1,lower.tail=FALSE)

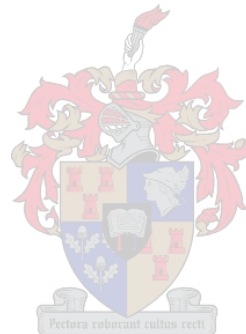
F2.age<-sum(FE[1:50]/diag(VFE1rob[1:50,1:50]))/sum(GMM[1:50]/diag(VGMM1[1:50,1:50]))
pF2age<-pf(F2.age,df1=50,df2=50,lower.tail=FALSE)
F2.year<-sum(FE[51:59]/diag(VFE1rob[51:59,51:59]))/sum(GMM[51:59]/diag(VGMM1[51:59,51:59]))
pF2year<-pf(F2.year,df1=9,df2=9,lower.tail=FALSE)

out<-
cbind(as.vector(GMM),as.vector(seGMM),as.vector(tGMM),as.vector(pGMM),as.vector(FE),as.vector(seFE),as.vector(tFE),a
s.vector(pFE), as.vector(F1.stat),as.vector(pF1))
dimnames(out)[[2]]<-list("GMM","std err","T","p-value","FE","std err","T","p-value","F","p-value")

```



```
#dimnames(out)[[1]]<-list("Age15","Age16", "Age17", "Age18", "Age19", "Age20", "Age21", "Age22", "Age23", "Age24",  
"Age25", "Age26", "Age27", "Age28", "Age29", "Age30", "Age31", "Age32", "Age33", "Age34", "Age35", "Age36", "Age37",  
"Age38", "Age39", "Age40", "Age41", "Age42", "Age43", "Age44", "Age45", "Age46", "Age47", "Age48", "Age49", "Age50",  
"Age51", "Age52", "Age53", "Age54", "Age55", "Age56", "Age57", "Age58", "Age59", "Age60", "Age61", "Age62", "Age63",  
"Age64", "Year1997", "Year1998", "Year1999", "Year2000", "Year2001", "Year2002", "Year2003", "Year2004", "Year2005")  
  
ageFE<-out[1:50,5]  
ageGMM<-out[1:50,1]  
  
list(out=out)  
  
detach(data4)  
  
}
```



Appendix 2: Summary of Cohort Sizes

Table 4 Cohort sizes: Birth Cohort Panel

Variable	Obs	Mean	Std. Dev.	Min	Max	Number of cohorts <100	Number of cohorts <200
Employment	561	1917.296	1138.345	289	5277	0	0
Unemployment	561	1205.159	879.983	26	3269	22	54
Participation	561	1917.296	1138.345	289	5277	0	0

Table 5 Cohort sizes: Birth and Population Group Cohort Panel

Variable	Black							White						
	Obs	Mean	Std. Dev.	Min	Max	Number of Cohorts <100	Number of Cohorts <200	Obs	Mean	Std. Dev.	Min	Max	Number of Cohorts <100	Number of Cohorts <200
Employment	561	1462.198	951.7025	201	4336	0	0	561	167.4118	66.56119	38	317	119	370
Unemployment	561	902.7647	696.5797	18	2656	36	87	561	112.1497	67.84535	0	282	264	485
Participation	561	1462.198	951.7025	201	4336	0	0	561	167.4118	66.56119	38	317	119	370

Table 6 Cohort sizes: Birth and Education Level Cohort Panel (Black)

Education Category	Variable	Obs	Mean	Std. Dev.	Min	Max	Number of cohorts <100	Number of cohorts <200
No education	Employment	561	153.2995	91.43504	18	486	191	403
	Unemployment	561	85.49376	55.09126	1	257	366	544
	Participation	561	153.2995	91.43504	18	486	191	403
Primary	Employment	561	431.8717	279.9622	69	2273	16	83
	Unemployment	561	264.2513	163.1222	8	760	105	232
	Participation	561	431.8717	279.9622	69	2273	16	83
Incomplete Secondary	Employment	561	593.8645	612.4319	28	3086	64	175
	Unemployment	561	317.4991	291.1732	2	1242	160	253
	Participation	561	593.8645	612.4319	28	3086	64	175
Matriculation	Employment	559	211.6583	266.8566	1	1058	294	361
	Unemployment	559	171.6691	225.8482	0	891	316	399
	Participation	559	211.6583	266.8566	1	1058	294	361
Some Tertiary	Employment	536	65.84701	61.86632	1	271	400	517
	Unemployment	536	60.54851	60.63142	0	265	416	522
	Participation	536	65.84701	61.86632	1	271	400	517

Appendix 3: Presentations in 3 Dimensions: Age, Year and Birth Cohort

Figure A 1 Unemployment rate by birth cohort and year

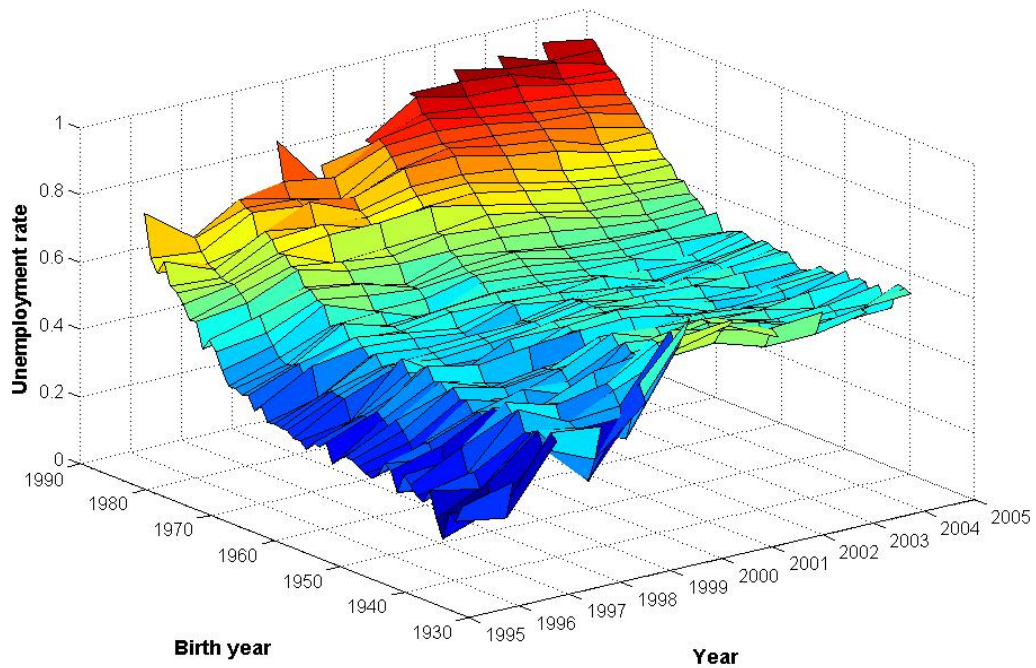


Figure A 2 Employment rate by birth cohort and year

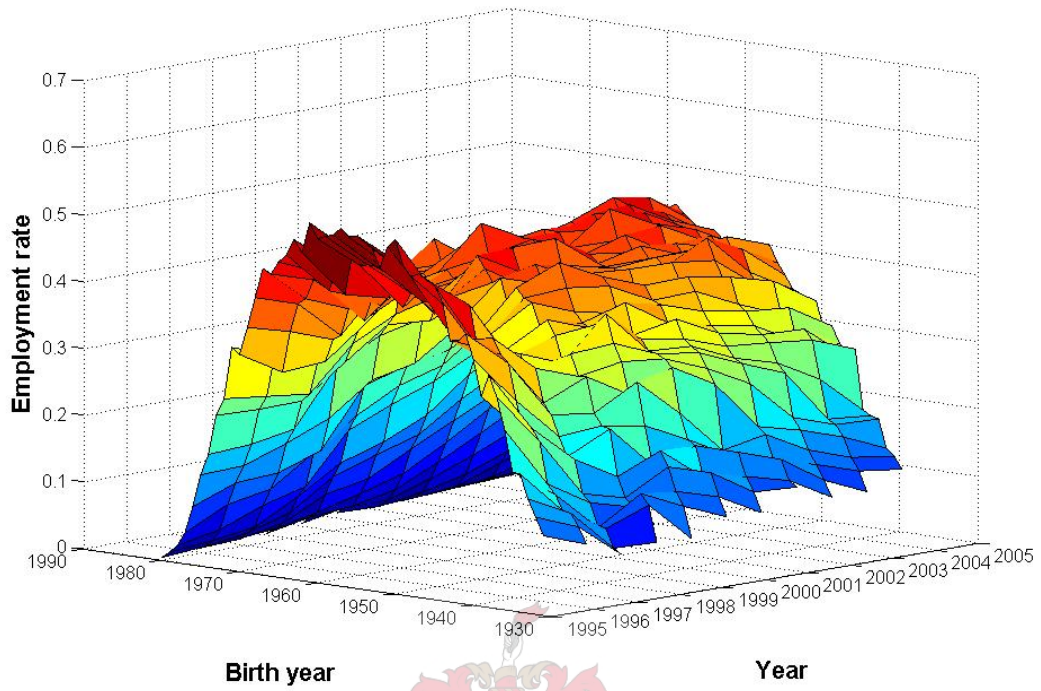


Figure A 3 Participation rate by birth cohort and year

