# NINTH ANNUAL CONFERENCE ON ECONOMETRIC MODELLING FOR AFRICA

**30 June**
**to**
**2 July 2004**

## Racial and Gender Wage Differentials in South Africa: What can Cohort Data tell?

Carola Gruen

# Racial and Gender Wage Differentials in South Africa: What can Cohort Data tell?

Carola Gruen
School of Economic and Business Sciences
University of the Witwatersrand

January 30, 2004

**Abstract**

Ten years after the first democratic elections, high poverty rates as well as a very unequal distribution of income and wealth are still striking features of the new South African society. Recent studies analysing the extent of income inequality show that since the end of Apartheid income distribution has at best not changed at all, but depending on the measure, might also have worsened in the last decade. The data used in these studies are repeated cross sections, thereby allowing a snapshot of the extent of income inequality at several points in time, but the analysis of temporal changes at an individual level is not possible. The following paper proceeds differently. By using subsequent years of the October Household Survey data it is possible to construct a synthetic panel. Preparing cross sectional data that way allows us to better utilise individual information and to address temporal developments also in the absence of genuine panel data. The paper focuses on gender and population group specific cohort wages to get a more detailed description of income inequality. Average earnings of birth cohorts of African and White workers employed full-time in formal sector jobs are followed over time and wage differentials as well as the mobility of cohort wages are studied in detail. A decomposition of African cohort wages into age, cohort, and year effects gives information about the existence of cohort effects. Results suggest that especially for young African women such generational trends may differ from the theoretical expectation.

JEL classification: D31, J31

Keywords: Cohort data, Gender and racial wage differentials, Generational trends

# 1 Introduction

Previous studies demonstrated that in the South African labour market the probability of finding regular employment as well as average wages differ greatly by race and gender (e.g. Kingdon and Knight, 2001; Rospabe, 2001a). Differences in measurable personal characteristics could only partly explain the observed gaps. The magnitude of explainable and unexplainable portions of particular wage differentials varies somewhat with the chosen specification of the wage regression and the decomposition technique applied. But regarding racial wage gaps between African and White workers, there is consensus that no declining trend of the unexplained part has developed in post-Apartheid South Africa until 1999 (e.g. Rospabe, 2002; Allanson and Atkins, 2001; Erichsen and Wakeford, 2001; Allanson, Atkins, and Hinks, 2000a).

Results like this are derived from a conventional decomposition of mean wage differentials. A linear model of wage determination is specified and the ordinary least squares estimator is used to predict individual wages. But only the average of these predictions is considered further on in most decomposition analyses, thereby disregarding the existing heterogeneity among the individuals.[1]

Studies examining South African labour market outcomes in the second half of the 1990's are often based on the October Household Survey data. These are cross sectional data and allow a very detailed description of e.g. the earnings situation of individual workers in the particular survey year (e.g. Rospabe, 2001b; Allanson, Atkins, and Hinks, 2000b). The analysis of subsequent years of the OHS data enables us to follow particular groups of people and to study changes of regional, occupational, or racial mean wages over time (e.g. Rospabe, 2002; Allanson and Atkins, 2001; Erichsen and Wakeford, 2001; Allanson, Atkins, and Hinks, 2000a). With cross sectional data at hand, it is however not possible to follow individual workers. Only panel or longitudinal data permit to view temporal changes at an individual level, i.e. whether people with relatively low income still belong to this income group in a later period (Baltagi, 1998; Deaton, 1997). To evaluate for example the success of anti-poverty strategies it is exactly such questions that have to be addressed.

The KwaZulu-Natal Income Dynamics Study (KIDS) partly offers an alternative to the OHS data. In the province of KwaZulu-Natal, African and Indian households which took part in the first South African national household survey in 1993, were re-surveyed in 1998.[2] The resulting panel was also subject to studies analysing employment and earnings mobility (e.g. Cichello, Fields, and Leibbrandt, 2002; Klasen and Woolard, 2002; Keswell, 2000).[3] However, as only households residing in that particular province, even though it is the most populous one, and only two population groups did participate in the survey, the sample size is rather small.

[1]An alternative to analyse earnings discrimination is suggested by Jenkins (1994). Following his approach, the discrimination measurement is not based on mean wages, but takes into consideration the complete distribution of predicted and reference wages.

[2]For further information on these data, see May, Carter, Haddad, and Maluccio (2000).

[3]Cichello, Fields, and Leibbrandt (2002) found that African workers in KwaZulu-Natal experienced rather volatile earnings. But not only the extent, also the direction of earnings movements was surprising. Low-income earners in 1993 had larger gains on average than those who started with relatively high earnings. Regarding the transition between formal and informal employment it turned out that movements out of regular employment were frequently accompanied by increases in real earnings.

Instead, I will use subsequent years of the OHS to construct cohort data (Deaton, 1985). Individuals who share particular characteristics (e.g. born in the same year) are pooled into cohorts and the means for each group are calculated. Applying this strategy to several survey years allows to follow (birth) cohorts over time and to build up a synthetic panel. That way, the issues of ignoring the heterogeneity among workers as well as the inability of studying the dynamic behaviour of individuals are at least partly tackled. Variables like the mean wage of African women can be split up into several age groups thereby revealing their contributions to the overall average in one particular year. As cohorts are tracked for various years, it is now also possible to watch their earnings mobility. The comparison of such within cohort changes will show whether or not young and old cohorts develop alongside similar paths. Finally, as different cohorts are observed at the same age, I will attempt to separate life-cycle from generational effects. This decomposition results in an age-earnings profile and allows to ascertain the existence and direction of cohort effects.

# 2 Constructing a Synthetic Panel

To demonstrate, how synthetic panels can be constructed and what potential problems are associated with this approach, it is helpful to begin at the individual level.[4]

Consider the linear model:

$$Y_{it} = X'_{it}\beta + \alpha_i + u_{it}, \qquad t = 1, ..., T \tag{1}$$

where subscript $i$ indicates individuals observed over $T$ periods. $X'_{it}$ is a set of explanatory variables, $\beta$ the set of parameters to be estimated, and the error terms with the commonly assumed properties are given by $u_{it}$. The $\alpha_i$'s represent unobserved individual effects which are constant over time, for example inherent ability or motivation. Such individual effects are likely to be correlated with the regressors.[5] As a standard approach the $\alpha_i$'s are treated like group specific constant terms. Since panel data observe the same individuals for more than one point in time, these invariant terms can be eliminated by a within or first difference transformation. The resulting differenced equation is then estimated by ordinary least squares (Greene, 2000).

Deaton (1985) suggests that to any linear individual relationship, as the one shown in equation (1), there exists a corresponding *cohort version*. With a series of independent cross sections being available, it is not possible to follow individuals or particular households, but to track cohorts over time.[6] Cohorts are formed among individuals who have one or more characteristics in common. Each individual belongs to one cohort only and this association is constant over time. Aggregating single information to cohort level and substituting individual observations by the cohort average result in the following model:

$$\bar{Y}_{ct} = \bar{X}'_{ct}\beta + \bar{\alpha}_{ct} + \bar{u}_{ct}, \qquad c = 1, ..., C, \qquad t = 1, ..., T \tag{2}$$

[4] The theoretical explanation mainly follows Verbeek and Nijman (1992) and Verbeek (1996). For further discussion see also Deaton (1985, 1997) and Baltagi (1995).

[5] If individual effects are assumed to be randomly distributed instead, a random effects approach is applicable, with $\alpha_i + u_{it}$ forming a composite error term.

[6] Early studies of life-cycle models were already based on cohort data but the conducted analysis was sometimes of a rather descriptive nature (e.g. Shorrocks, 1975).

where, for instance, $\bar{Y}_{ct}$ is the average value of all observed $Y_{it}$'s in cohort $c$ at time $t$. Regarding individual fixed effects, the aggregation to cohort level leads theoretically to cohort fixed effects, if a constant (cohort) population is assumed.[7] In practice, however, the average is taken over the surveyed cohort members only. Since for each period different individuals are observed, $\bar{\alpha}_{ct}$ is not constant over time, "because it is the average of the fixed effects of different [...] [individuals] in each year" (Deaton, 1997, p. 122). Thus, the $\bar{\alpha}_{ct}$'s cannot be cancelled out by a transformation similar to the ones applicable to genuine panel data.

The time variation of cohort effects is, however, negligible, if the number of observations per cohort is large. In that case, the model changes to:

$$\bar{Y}_{ct} = \bar{X'}_{ct}\beta + \bar{\alpha}_c + \bar{u}_{ct}, \qquad c = 1, ..., C, \qquad t = 1, ..., T. \tag{3}$$

The resulting pseudo panel consists of $T$ observations on $C$ cohorts and the parameter vector $\beta$ could be determined using the standard within estimator $\hat{\beta}_w$.

Deaton (1985) correctly points to an additional measurement error problem. Dependent and independent variables are measured by the observed cohort means $\bar{Y}_{ct}$ and $\bar{X'}_{ct}$ which are error-ridden estimators of the unobservable population cohort means $Y^*_{ct}$ and $X^{*'}_{ct}$. The measurement error on the independent variables causes the estimator to be biased toward zero. By applying errors-in-variables techniques it is possible to correct for this. Since the data are available on the individual level, both cohort averages and associated standard errors can be estimated. The estimated standard errors correspond to the variance due to measurement error which then has to be subtracted from the classical estimator to yield a consistent estimator, the later on so-called Deaton's errors-in-variables estimator $\hat{\beta}_D$.[8]

Verbeek (1996) examines the consistency properties of both the within estimator $\hat{\beta}_w$ and Deaton's errors-in-variables estimator $\hat{\beta}_D$ with respect to the total number of observations $N$, number of cohorts $C$, number of observations per cohort $n_c$, and periods $T$. He concludes that the cohort size $n_c$ is crucial for the consistency of the within estimator. If $n_c$ tends to infinity, $\hat{\beta}_w$ and $\hat{\beta}_D$ become equivalents. This finding supports the approach of many applied papers which argue that if cohort sizes are relatively large, it is possible to ignore the measurement error and use the standard within estimator.[9] But even relatively large cohort sizes may not be sufficient to reduce the bias significantly as shown by

[7] This assumption is necessary as consecutive surveys are used to generate random samples from the same underlying population. In most applications, it is reasonable to suppose an invariant population. The literature frequently discusses two circumstances that may lead to a violation of this assumption. Firstly, if an economy is subject to e.g. substantial migration or death rates, the population structure alters over the years. Secondly, when working with household instead of individual data, cohorts are often defined by the age of the household head. However, being head of household is not a constant characteristic but households get reorganised in case of marriage, divorce, or if old people dissolve their own household to live with their children (Deaton, 1997; Moffitt, 1993).

[8] An alternative approach that allows to estimate also dynamic models from a time series of cross-sections was introduced by Moffitt (1993). He disregards the error-in-variables problem and demonstrates that repeated cross sectional data can be used to consistently estimate models with lagged endogenous variables. Since the samples are independently drawn, observed changes in cohort averages are also independently measured. It is therefore possible to use changes from earlier years as instruments (see also Deaton, 1997; Verbeek, 1996). For an application of this instrumental variables procedure see for example Blundell, Browning, and Meghir (1994).

[9] See for example Browning, Deaton, and Irish (1985) with cohort sizes of 190, Deaton and Paxson (1994a) where $n_c$ lies between 150 and 400, Blundell, Meghir, and Neves (1993) who on average have 360 observations per cohort, or Jappelli (1999) grouping up to 700 individual observations.

Verbeek and Nijman (1992):[10]

$$\omega = plim_{C \to \infty} \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{X'}_{ct} - X^{*'}_{ct})^2 = \frac{1}{n_c} \sigma_v^2. \tag{4}$$

To minimise the measurement error variance $\omega$ in $\bar{X'}_{ct}$, clearly, a large number of observations per cohort is necessary. However, the way individuals are aggregated is also important as it influences $\sigma_v^2$, the within cohort variance. Cohorts should be constructed such that within-variation is small and between-variation is large. In other words: "[...] individuals within each cohort should be as 'homogeneous' as possible, while those from different cohorts should be as 'heterogeneous' as possible" (Verbeek, 1996, p. 284). In addition, inspecting the standard errors of the cohort means ensures that regression results are not dominated by the way of sampling (Deaton, 1997).

The number of observations per cohort does not only influence the magnitude of the measurement error, but also determines the size of the pseudo panel and thus the variance of the within estimator. An optimal choice of cohort sizes will therefore take into account consequences for both the bias arising from measurement error and the variance of the estimator. Verbeek and Nijman (1992) examine the magnitude of these two opposite effects. They show analytically that an increase in $n_c$ finally results in an increase in the variance of $\hat{\beta}_w$ and confirm this finding by an empirical test.[11] The effect of a smaller variance if cohort averages are estimated more precisely is more than offset by an increase in the variance if the estimation has to be based on a smaller total number of observations. The existing trade-off between cohort size and the number of cohorts is frequently referred to when discussing the cohort design. In case of the errors-in-variables estimator it is possible to choose an optimal cohort size, but when relying on the standard within estimator, one basically has to weigh up the bias with the variance (Verbeek, 1996).

After having explained how pseudo panels can be constructed, it is reasonable to discuss the usefulness of such strategy. If long series of cross sectional data are available, this method enables us to study the dynamics of particular age groups. Various estimators for static and dynamic models have been developed to yield consistent results when applied to cohort data (see for example Verbeek and Vella, 2000; Collado, 1997; Moffitt, 1993; Verbeek and Nijman, 1993; Deaton, 1985). From a theoretical point of view, synthetic panels may be even preferred to genuine panels. The problem of attrition due to mortality, refusal, or mobility let panel data become less representative over the years. Since cross sectional data are based on a newly drawn sample each time, they better fulfill the criterion of representativeness.[12]

From an econometric standpoint, working with grouped data has two major consequences. Parameter estimates are less efficient, since data aggregation leads to a loss of

[10] For the sake of simplicity, it is assumed that cohort sizes are equal (i.e. $n_c = N/C$). Otherwise, observations have to be weighted by the square root of the cohort size first to obtain a homoscedastic variance (Greene, 1997; Verbeek and Nijman, 1992).

[11] As the derivation of this result is rather complex, I refer to Verbeek and Nijman (1992) for a more detailed discussion. In their empirical part, they analyse food expenditures of Dutch households and come to the conclusion that "fairly large cohort sizes (100, 200 individuals) are needed to validly ignore the cohort nature of the data" (Verbeek and Nijman, 1992, p. 20).

[12] Similar problems may of course arise if the sample design is changed over time or if not all population groups have the same probability of being selected into the sample (Deaton, 1997).

information.[13] The fit of grouped data regressions is considerably higher, sometimes close to one (Greene, 1997). If the number of observations per group differs considerably, heteroscedastic error terms will occur in addition. Weighting the data by the square root of cohort sizes ensures that the error terms are homoscedastic and thus all assumptions of the classical linear regression model are met (Deaton, 1985; Greene, 1997).

In the following individual data of African and White workers will be aggregated to form (birth) cohorts of full-time formal sector workers.[14] The resulting data structure allows to examine wages and wage differentials at a disaggregated level. As the cohorts are tracked over time, it is also possible to describe a particular wage path per cohort. Such wage paths may develop differently for young and old cohorts. Lining up cohort wages by age shows a particular pattern which can be split up into several components. That way, it is possible to detect the contribution of age, year of birth, and common shocks to the observed cohort-earnings curve.

# 3 Cohort Wages for African and White Workers

The independent cross sectional data for the years 1995, 1997, and 1999 of the October Household Survey are used to construct a synthetic panel. The variable of main interest is (the logarithm of) the real hourly wage earned by full-time workers employed in the formal sector.[15] Aggregation could be based on measures of central tendency such as the mean and median or particular percentiles. To ascertain, whether the data used here contain extreme values, cohort averages were calculated applying both the mean and median of real hourly wages. Cohorts are defined by gender, race, and age. The mean wage exceeded the median wage in each cohort, indicating that the within cohort distribution of real hourly wages is skewed to the right. In some cohorts few extreme individual values caused the mean to deviate substantially from the median. These outliers were excluded and the following analysis uses the arithmetic mean of the logarithm of real hourly wages as cohort average.[16]

The definition of cohorts and corresponding cell sizes are reported in Tables 1 and 2. Aggregating individual data according to race and gender and applying a five-year age band results in six cohorts per population group covering workers aged between 20 and 49 in 1995 (Table 1).[17] For example, the first cohort pools all formal sector full-time workers aged 20 to 24 in 1995. In 1999, workers at the age of 24 to 28 belong to this cohort. For all population groups cohort sizes first increase with age, peak out at the

[13]On the other hand, in the context of measurement error grouped variables are recognised as an instrument since averaging may reduce or cancel out measurement errors. The loss in efficiency would then be smaller (Kennedy, 1998).

[14]The analysis is restricted to full-time formal sector employees to at least partly assure that wages earned in comparable employment status are considered only. For a more detailed description of the data preparation see Grün (2003).

[15]The 1997 and 1999 wages were deflated to the base year 1995 using the consumer price index (Stats SA, 2001). In the following, *workers* refer to people working full-time in a formal sector job.

[16]The least differences between mean and median wages were found among White female workers, whereas among White men aged 35 to 39 in 1995 seven outliers have been excluded. Regarding African workers, real hourly wages are much more dispersed for both gender. Unreasonably high wages were found in four cohorts and altogether nine observations of African workers were precluded from further analysis.

[17]The number of observations available for White women limits the analysis to these age groups.

middle cohorts covering workers in their thirties in 1995, and decline thereafter again. The greatest concessions regarding reasonable cell sizes had to be made for White female workers, where the number of observations per cohort is sometimes less than 100. With respect to African workers, cohort sizes are fairly large. To get a more detailed picture for this race, a second set of cohorts has been constructed applying a two-year age band. As shown in Table 2, this results in 14 cohorts available for analysis. Descriptive statistics for African workers and the decomposition analysis are based upon these data.

Before turning to the results it should be pointed out explicitly that the following analysis is more of an explorative nature. The synthetic panel is based on three subsequent cross sections and altogether, cohorts are followed over a time span of five years only. The scope to observe for example different cohorts at the same age is thus rather limited and the obtained results do not represent consolidated findings. Regardless of these limitations, the analysis allows interesting insights and can at least hint to differences that exist between different groups of workers.

## 3.1 Summary Statistics Based on Cohorts

Figure 1 gives a first look at the logarithm of real hourly wage by cohort for the different population groups.[18] Each connected line represents the mean wage for one cohort in 1995, 1997, and 1999. Wages earned by African male and female workers are given in the upper panel. Since different cohorts are observed at the same age, lines do overlap. When tracing the wage level across different cohorts it becomes obvious that wages increase with age but at a decreasing rate. Among older cohorts of women mean wages even seem to decline again.[19] Facing the age-group specific averages of one year with the overall mean in that period provides insights into the composition of the latter.[20] For African men, the comparison confirms what the graph had already suggested. In 1995, the mean wages of the first three cohorts, which are incidentally among the lowest observed in the sample, pull down the overall average to 1.73. Overall averages for the subsequent years amount to circa 1.8 as the young cohorts realised considerable increases and older cohorts enjoyed relatively stable earnings for the total time span. Regarding females, the rather heterogeneous picture of cohort wages is striking. To find out whether the various within cohort changes are statistically significant they will be discussed separately. For the moment, it suffices to note that African women had their highest overall mean wage of about 1.83 in 1995 due to the relatively high earnings of the middle cohorts in that year. As these cohorts suffered on average from considerable wage losses in the following years, overall means dropped to 1.71 in 1997 and 1.67 in 1999, respectively.

The lower panel presents cohort wages for White male and female workers. Since for this race the five-year interval has been applied, it is not possible to observe different cohorts at the same age. For both gender the data point to higher wage levels as workers belong to older cohorts. Regarding the total time span, men realised on average somewhat bigger increases. Within cohort trends suggest that White women moved along a relatively smooth path, while males had on average a particularly bad year in 1997. This impression is confirmed when considering the overall averages. In 1997, White men realised the

[18]In the following, *wage* refers to the logarithm of real wage earned per hour.

[19]Such concave age-earnings paths are discussed in relation to the theory of human capital. See section 4 for further discussion.

[20]I refrained from adding the overall means for the three years to the graphs to keep the pictures clear.

lowest mean of 3.03 and experienced average wages of about 3.15 in the adjacent years. Only White women enjoyed on average constantly increasing wages as the overall mean amounted to 2.73, 2.76, and 2.83 in 1995, 1997, and 1999.

Inspecting related socioeconomic characteristics from a cohort perspective is instructive as well. A detailed look on variables determining wage levels and employment propensity is given in Figures 2 and 3 for African and White workers, respectively.[21] The first panel in Figure 2 shows the number of years of schooling completed by cohort, separated for African men and women. With the exception of the youngest cohort in 1999, the level of education was on average higher among African women employed in a formal sector job than among African men in this sector. Although the differences are rather volatile, the gap becomes somewhat smaller for younger cohorts. Moving from younger to older cohorts, the number of schooling years follows a downward trend for both gender. One would expect older cohorts to be less educated than younger ones, but the level of formal education within cohorts should be relatively constant.[22] Since the distinct decline for middle and older cohorts is also revealed when looking at broader samples like labour force participants and working age population, it cannot be attributed to an education-based selection out of formal sector employment. It would appear that the decline might be due to the changing way of reporting educational levels.[23] This assumption, however, is not confirmed when turning to White workers (Figure 3, first panel). Here, within cohort educational levels do not indicate any particular trend and also across cohorts changes only happen on a small scale.

Following the number of children living in the workers' household across cohorts points to different patterns for Africans and Whites (Figures 2 and 3, second panel). For the latter the number of children at first increases with age, reaches its maximum around the age of 35-40, and decreases thereafter. Although the observations for African workers are noisier, it becomes clear that there is no such inverted U relationship between age and the number of children living in the household.[24] A maximum number can still be made out for both gender, but a downward trend is discernable for females only. This decline, however, comes to an end as women grow older and finally turns back into an increase.[25] Adapting results of Klasen and Woolard (2000), who examine household formation in the context of unemployment, the re-increasing number of children might be related to a better access to resources when being attached to a worker's household.

The last panel in Figures 2 and 3 shows for each cohort the proportion of cohort members acting as head of household. Regarding men, this proportion is steadily increasing both within and between cohorts and comes close to 100 per cent for the oldest cohorts. Within single cohorts of African women the share also rises considerably between 1995 and 1999. The proportion of household heads among African female workers aged 44 and above in 1995 lies between 40 and 60 per cent. Among White peers the share is only half as much.

[21] In the following, the focus will be on particular findings exposed only when examining cohort data. For a general discussion of these variables as well as their relevance for finding employment and wage setting, see for example Grün (2003).

[22] It may of course vary in a random way, as every year different individuals are grouped together.

[23] Respondents were given a more detailed list of school grades and degrees in later years.

[24] According to the data, the number of children living in African households peaked in almost every cohort in 1997. It is, however, hard to tell what year(s) might have caused this somewhat peculiar finding.

[25] Admittedly, this course is hardly visible when looking at the graph based on the workers' sample, but shows up a lot clearer when viewing for example labour force participants.

This last example especially points out advantages when working with synthetic panel data. Summary statistics obtained from cross sectional data would have indicated that the proportion of female headed household among workers is nearly twice as high for Africans than for Whites (e.g. Grün, 2003). Considering various age groups at one particular year would have revealed differences existing *between* younger and older individuals. In the absence of individual panel data, only cohort data constructed from a time series of cross sections facilitate to study also temporal developments *within* particular age groups.

## 3.2 Cohort Specific Wage Differentials

The upper panel of Figure 4 plots cohort wages for White and African men, the lower panel cohort wages for female workers of both races.[26] The racial wage hierarchy known from analysing cross sectional data continues at the disaggregated level: cohort wages for White workers are always above the level of the corresponding African cohorts. Comparing the upper and lower panel of Figure 4, the gaps emerging between cohort wages of male workers are always larger than those that exist between female workers. Within cohort wages develop quite differently over time and will be examined separately in the next section. Abstracting from within cohort variation, cohort earnings of White and African men follow similar paths as workers age. As a result, the gap between wages of younger cohorts is commensurable to the one appearing for older cohorts. Regarding females, the racial gap between cohort wages tend to get larger among older cohorts. But this tendency turns out to be not statistically significant, as the confidence intervals become larger for these group of workers.

Figure 5 shows similar graphs for cohort specific wage differentials between male and female workers holding race constant. The gaps shown here are substantially smaller than the racial differences. Focussing African workers depicted in the upper panel, the overall evolution of wages across cohorts is rather similar, except for the oldest cohort. The wage gap between men and women aged 45-49 in 1995 broadens noticeably and becomes finally significant in a statistical sense in 1999. For the year 1995, previous studies relying on cross sectional data found out that wages earned by African females are on average higher than wages of African men (see Grün, 2003; Hinks, 2002; Erichsen and Wakeford, 2001). The breakdown into several age groups reveals that especially younger cohorts contribute to this finding. But even if overlapping confidence bands indicate that in nearly all cases the negative wage differential is not statistically significant it remains an unusual outcome.

As regards White workers, the majority of cohort specific gender wage differentials turns out to be statistically significant. Looking at younger cohorts of both gender, they, however, experience comparable wage levels as well as growth rates resulting in only small and mostly not significant wage differences. Among workers who are in their thirties in 1995 the gender wage gap increases considerably, as female wages now diverge substantially from the men's wage level. Relating this finding to the discussion on the number of children living in the workers' household suggests that on average women at this age cut back their labour market activities to raise children. Moving on to older cohorts the gender gap narrows and becomes statistically insignificant again, as men's wages do on average no longer experience positive growth rates but approach the wage level of women.

[26] To compare the same birth cohorts of African and White workers, cohorts for the former are now also based on the five-year age band.

To conclude this part of the analysis, it should be emphasised again that wage differentials observed between groups of workers do not automatically point to unequal treatment in the sense of discrimination. The unexplained portion of wage gaps when averaging over all age groups sometimes assumes alarming proportions (e.g. Grün, 2003; Rospabe, 2001a, 2002; Allanson, Atkins, and Hinks, 2000b). One would expect to see a similar pattern at a disaggregated level. Splitting the workers sample into several age cohorts reveals that cohort specific gaps are relatively constant across cohorts or increase and decrease as workers age. The question then arises, whether similar gaps between age cohorts, as those between White and African men, correspond to similar unexplained components or rather mask changing magnitudes of the explainable and unexplainable portions of the wage gaps. However, this study does not intend to decompose wage differentials to answer this question as the number of observations in the synthetic panel is too small to estimate wage regressions. Furthermore, variables used in panel estimations should exhibit a certain degree of random variation over time. But regressors like the level of formal education and age either do not change over time or vary in a systematic way.

## 3.3 The Dynamics of Cohort Wages

The structure of the synthetic panel explicitly allows to follow cohorts over time. Although the discussion in previous sections already mentioned wage growth within particular birth cohorts, it will be addressed in detail now to see whether for example young and old cohorts differ in their experienced earnings mobility. As only three points in time covering a total time span of five years are considered, the analysis of within cohort growth will be restricted to a rather short period, but can still point to different tendencies for different cohorts.

Figure 4 plots cohort wages for all four groups of workers without intersection. Solid lines in the first graph represent cohort wages for White men, that group of workers who earned on average the highest wage at any given age. When considering the total time span, only the youngest cohort realised a wage increase that also turns out to be statistically significant. Moreover, the gain is in absolute numbers the biggest observed for the total sample. As indicated by the confidence intervals, White male workers aged between 25 and 34 in 1995 experienced relatively stable earnings, in contrast to older cohorts who on average realised losses in real wages between 1995 and 1999.[27]

Only middle and older cohorts among African female workers had to bear similar or even larger declines in earnings in absolute (log) terms (see dashed lines in the lower panel). The already received impression of a very heterogeneous group of workers continues when looking at within cohort changes. Cohort wages jump up and down quite considerably. As regards the youngest cohort this results in no significant trend for the total period. It is thus the only population group of workers aged 20 to 24 in 1995 that cannot realise significant wage increases over the total time span covered here.[28]

[27]The 95 per cent confidence bands for the last two cohorts do slightly overlap. But the difference between 1995 and 1999 for the two oldest cohorts of White male workers as well as similar marginal cases among other population groups are significant at an interval of 10 per cent.

[28]This rather pessimistic result persists when examining cohorts defined by the two-year age interval. Then, the first three cohorts do not experience significant wage changes between 1995 and 1999.

Within cohort changes of the two remaining groups of workers follow similar patterns. The first two and three cohorts of African men and White women respectively enjoy significant wage increases between 1995 and 1999. Regarding older cohorts, the first impression gained from the descriptives in Figure 1 is confirmed: wages within cohorts remained relatively constant for the period considered. Concerning White women, however, the insignificance of wage differences might partly be due to increasing variance as the number of observations per cohort becomes very small.

Already the limited time frame of five years clearly shows that within cohort wages develop differently for younger and older workers belonging to the same population group. Younger cohorts did realise statistically significant wage increases, except for African women, whereas older cohorts either faced relatively unchanged wage levels or had to cope with real wage losses. Comparing within cohort growth of workers who belong to different groups of the population the analysis suggests that cohorts of young White workers benefitted most from wage increases. Unlike African women, who on average were found to be in an inferior situation, as young cohorts did not realise wage increases but already middle aged workers suffered from declining mean wages.[29] However, to yield consolidated findings on within cohort trends, a greater number of periods $T$ is needed.

# 4 Determining Age, Cohort and Year Effects

## 4.1 General Remarks

Economic theory suggests for a variety of (socio-)economic quantities a particular pattern evolving over an individual's working life or total life time. For example, Modigliani's life cycle hypothesis of saving argues that individuals, while working, will save a certain fraction of their income thereby accumulating wealth which is continuously reduced once they are retired. As a result, the relationship between age and wealth will be humped-shaped (Modigliani, 1986). As economies grow, variables like savings, incomes, and wealth are also subject to secular trends. Alongside economic development savings and incomes are growing, whereas for example household sizes tend to decline (Deaton, 1997). In the absence of longitudinal data synthetic panel data have been frequently used to verify various age-related profiles and to disentangle the age effect from the generational or cohort effect.[30]

The relation between age and earnings also follows a distinct pattern. Young workers start with relatively low wages but the average wage level increases as workers grow older. Since the increases diminish over time the overall curve turns out to be concave. A theoretical framework is given by the human capital model (see for example Becker, 1993;

[29]This does of course not imply that once contemporarily young cohorts are aged they will pass through the same path of wages as the one described for currently old cohorts. See also the next section.

[30]For example, Jappelli (1999) tested the age-wealth profile for Italy, Deaton and Paxson (1994b) examined levels of income, consumption, and saving in Taiwan using cohort data. Pseudo panels have also been used to study the relation between consumption inequality and age in the UK, USA, Taiwan, and Japan (Attanasio, Berloffa, Blundell, and Preston, 2002; Ohtake and Saito, 1998; Deaton and Paxson, 1994a).

Mincer, 1962).[31] The model designs a particular wage path over the working life cycle of individuals. Empirical tests whether concave age-earnings profiles exist are ideally based on longitudinal data. Profiles derived from cross sectional data can appear differently because of secular trends toward higher education and occupational or life-cycle employment changes, for example.[32] Unless such cohort effects can be neglected or do not exist, as in case of a stationary economy, age-earnings profiles obtained from time series and cross sectional data are identical (Becker, 1993).

Table 3 brings together real hourly wages (now not in log form) for various groups of workers at two different points in time. To be able to compare the income of a particular age group in 1995 with the appropriate cohort's income in 1999, a four-year age interval has been chosen. The *cross sectional* profiles for all groups of workers are humped-shaped: wages first rise with age, reach a maximum, and decline thereafter. Looking at the data this way also suggests that highly skilled workers realise their maximum earnings later than less skilled workers as the peak in earnings for Africans, who on average work in occupations requiring less skills, happens at an earlier age than for Whites. However, the underlying statistic is not appropriate for such a conclusion. Cross sectional data observe workers of different age at one point in time and the resulting age-earnings profile is composed of wages earned by people with different life time earnings. Therefore, the earlier peak for less skilled workers might be a similarly wrong implication as the frequently referred to overstated downturn in earnings for older workers because cohort effects operating towards higher life time incomes for younger workers (regardless the educational or skill level) cannot be taken into account when analysing single cross sections (Deaton, 1997; Becker, 1993).

Turning again to Table 3 and comparing the cohorts' income in 1999 with the cross sectional income of the corresponding age group in 1995 makes clear that the two approaches yield different results. But only among older cohorts of White female workers does the cross sectional profile show a more pronounced downturn in earnings. Group specific wages of other workers in 1995 are either similar to or even above the corresponding cohort level in 1999. This somewhat surprising result suggests two things: firstly, as differences between cross sectional and cohort data exist it is important to take generational trends into account and secondly, those trends might be different from the expectation.

A third component that can be identified when decomposing earnings is a time effect which captures macroeconomic shocks affecting all cohorts in the same way. Equation (5) illustrates the functional form of such a decomposition.[33]

$$\bar{Y}_{ct} = \alpha + age_{ct} + cohort_c + time_t + u_{ct}, \qquad c = 1, ..., C, \qquad t = 1, ..., T. \tag{5}$$

The average of the logarithm of real hourly wages of cohort $c$ at time $t$, $\bar{Y}_{ct}$, can be separated into an age effect reflecting the typical profile evolving as workers grow older, a

[31] According to the theory, differences in earnings (and other labour market outcomes) can be explained by different amounts of human capital. Individuals want to maximise their life time income which is, among other things, subject to the level of human capital acquired. By choosing an optimal amount of both formal education and on-the-job training workers can increase their productivity and maximise their earnings (see also Blau, Ferber, and Winkler, 1998).

[32] For a brief review of analyses either based on cross sectional or longitudinal data conducted in the fields of labour economics and forensic economics see for example Gohmann, McCrickard, and Slesnick (1998) and Rodgers, Brookshire, and Thornton (1996).

[33] As it will be applied here, the decomposition is used as a descriptive device only. Regarding work on savings, consumption or wealth the decomposition into age, cohort, and time effects also allows to validate economic models, for example the life cycle hypothesis (e.g. Attanasio, 1997).

cohort effect mirroring secular trends, and a time effect absorbing aggregate shocks as for example business cycle effects (Deaton, 1997). Age control variables are easily specified, cohorts can be conveniently labelled by age at time $t = 1$, and year dummies are often implemented to pick up time effects. Equation (5) rules out any interaction between the single components thereby assuming the same age profile for all cohorts.[34] The cohort effect then indicates different positions of that age profile. If variables controlling for age, cohort, and time effects enter equation (5) in an unrestricted way, this will lead to an identification problem because of the linear relation between these three effects. For example, as soon as age and year are given, the corresponding cohort (defined by year of birth) is known as well and it is not possible to identify all effects simultaneously.

There are several ways to overcome this problem. If one is willing to assume certainty, which in turn means no unexpected common shocks, there is no need to include any control variables with respect to time.[35] Many applied papers, however, take the existence of macroeconomic influences into account and impose an additional restriction. Following Deaton and Paxson (1994b), all trends observable in the data can be attributed to age and cohort effects, if one assumes that the time effect has a zero mean and is not following any trend.[36] In the context discussed here, it seems reasonable to ascribe wage increases to age and cohort effects and to assume that cyclical fluctuations are zero in the long-run (Deaton, 1997). Studies decomposing earnings (e.g. Johnson and Stafford, 1974), savings (e.g. Kitamura, 2001; Jappelli and Modigliani, 1998), or consumption profiles (e.g. Bardazzi, 2000; Ohtake and Saito, 1998) applied this normalisation strategy or adopted slightly modified versions (e.g. Jappelli, 1999).[37]

## 4.2 Decomposing the Earnings of Africans

In the following, average earnings for the cohorts of African female and male workers shown in Table 2 will be subject to a decomposition into age, cohort, and year effects.[38] However, it should be pointed out from the beginning that this empirical investigation will probably encounter difficulties. As only three survey years are used to construct the synthetic panel and furthermore a two-year age interval has been chosen, successive cohorts are observed at the same age for a short period only. Thus, overlapping is extremely limited and it will be hard to distinguish between trends and transitory shocks (Deaton, 1997). Obtained results should therefore be interpreted cautiously.

The upper limit of each age interval specifies the age variable. Cohorts are consecutively numbered so that higher numbers correspond to older cohorts. One year dummy

[34]In principle, it would be possible to include interaction terms to allow, for example, in the presence of macroeconomic shocks younger cohorts to adjust savings differently than older ones (e.g. Attanasio and Weber, 1994). However, such extensions are not followed up here because, given the small size of the synthetic panel, the number of right-hand side variables should be kept at a minimum.

[35]That way, one would also forgo to capture any non-random influences not correlated with age or cohort, like measurement errors (Jappelli, 1999).

[36]To model orthogonality and zero mean, the standard year dummies $time_t$ have to be redefined to equal the following expression: $time_t^* = time_t - [(t-1)time_2 - (t-2)time_1]$. Finally, only $T - 2$ year dummies enter the regression to facilitate identification (Deaton, 1997; Deaton and Paxson, 1994a).

[37]A general discussion on normalisation and associated pitfalls can be found in Heckman and Robb (1985) who conclude "the real problem is finding [...] better explanatory variables and sharper behavioral models" (Heckman and Robb, 1985, p. 148).

[38]Since for Whites different cohorts are never observed at the same age, it is not possible to decompose the earnings of these workers in a similar way.

has to be redefined that time effects average to zero and are orthogonal to any trends. All observations have been weighted by the square root of cohort sizes $\sqrt{n_c}$ to accommodate the heteroscedastic nature of the aggregated data. Earnings can be regressed on both polynomials and a set of dummy variables. Table 4 shows the estimation results separated for African men and women. Columns labelled by (1) show the specification which fitted the data of male and female workers best under the consideration that the number of the remaining degrees of freedom is still acceptable. Regarding men, age and cohort polynomials are included, while for women, a combination of a third-order age polynomial and cohort dummy variables was chosen.[39]

Figures 6 and 7 provide a graphical illustration of the estimated effects according to specification (1). The top right-hand panels display the time effect that has been calculated from the restricted year dummy. Only for men, the coefficient is significant but compared to the other effects only small in magnitude. Given the short time span covered, it is difficult to interpret the time effect in its theoretical sense of capturing long term, aggregated shocks. But it suggests that in 1995 wages were on average higher even though age and cohort effects were controlled for.[40]

The bottom panels show age and cohort effects. Although the relationship between earnings and age is humped-shaped for both gender, the exact shape differs considerably between the two. Still appropriate for both groups, average wages peak around a relatively early age of thirty.[41] The men's profile increases steeply at first and declines moderately after the maximum. By contrast, the upward trend of the females' age effect is less pronounced among young workers, but as women grow older they face a sizable downturn in earnings. Wages realised by female workers aged forty correspond on average to those observed for the youngest. With advancing age the drop in earnings continues. This substantial downward trend is partly offset by cohort effects that are increasing with cohort age. Contrary to the theory outlined above, 'secular trends' are in favour of older cohorts. As regards men, the estimation result is less controversial since no significant cohort effect could be determined. As mentioned before, the short time span may hamper a correct separation of short term deviations and long term trends. Otherwise, the theoretical argument that generational effects are in support of younger cohorts also hinges on the economy's growth rate (Becker, 1993). For the period 1985-1999, the average annual growth rate of South Africa's GDP per capita amounted to -1.0 per cent (WDI, 2002). Therefore, we no longer would expect to see cohort effects favouring younger workers. In addition, if incumbent workers would not suffer from wage reduction given high and rising levels of unemployment, but young entrants had to agree upon lower initial payments, this could culminate in a positive cohort effect. This hypothesis, however, can be backed up only partially. Studies analysing the long term trend of African real wages in the formal sector document a certain rigidity of mean wages despite rising unemployment (e.g. Fedderke and Mariotti, 2002; Fallon and Lucas, 1998).[42] But whether young and older

[39] In all specifications both dummy variables and polynomials of different orders were tested to control for age and cohort. Both approaches yielded similar results and thus statistical parameters were decisive (i.e. R squared, degrees of freedom left, t-statistics, F-Tests).

[40] Specifications using normal year dummies but leaving out either age or cohort controls confirm this interpretation.

[41] Please note that the dependent variable is expressed in logarithms. Running the decomposition on antilog values leads to later peaks for both groups. However, all control variables for age and cohort in the female specification become insignificant.

[42] Fallon and Lucas (1998) conclude that other factors (e.g. African trade unionism) counteracted the exerted downward pressure on wages arising from high unemployment rates.

cohorts of workers contributed differently to this outcome is hard to answer as officially released numbers of entry wages or wage levels by age groups were not available.

The regression results prove to be relatively insensitive to modifications of the model. In specification (2) a control variable for educational attainment has been added. Figures 8 and 9 illustrate the results. In accordance with the theory of human capital, higher levels of formal education result in higher wages. For both gender, the age effect follows a compressed but similar path to the one obtained in the previous estimation.[43] The positive correlation between average wages and age in 1995 as a cohort measure still prevails for women and becomes slightly significant for men as well.

The last specification shown here relaxes the assumption of appropriate cohort sizes and considers all workers aged between 15 and 60 in 1995. That way, it is possible to check whether the results are driven by the chosen age span ignoring older workers in particular. But again, the overall picture does not change much (see also Figures 10 and 11). The age profile for men is concave, for women it is clearly humped-shaped. Almost all dummy variables controlling for different cohorts of women have positive and significant coefficients. For men, the effect becomes insignificant again.[44]

An assessment of the results obtained here turns out to be difficult since studies addressing similar questions about cohort wages in the South African labour market could not be found. There is some literature on earnings and household income mobility of Africans using the KIDS panel data (see for example Cichello, Fields, and Leibbrandt, 2002; Klasen and Woolard, 2002). Both studies conclude that in KwaZulu-Natal mobility between 1993 and 1998 was high and that changes in labour market status as well as movements between the formal and informal sector largely contributed to this. However, given the relatively small sample size and the fact that individuals were interviewed only twice, a cohort analysis with this data could hardly be more instructive.

The number of periods $T$ is crucial in any cohort analysis as it determines for how long cohorts can be followed over time and thus how many cohorts are observed at the same age. The previous analysis tracked cohorts over the short period of five years. It thereby only allows a direct comparison between the average wage of a given cohort with mean wages for somewhat younger or older workers. In other words, although belonging to different birth cohorts the particular groups of workers might not really have encountered different secular trends that could in turn be detected by any decomposition. To identify such effects it is necessary to follow cohorts over an extended period of time. Only then, the observed paths may non-randomly differ and a decomposition can reveal to what extent age and cohort effects have contributed to this outcome.

# 5 Concluding Remarks

The aim of this work was to go beyond a review of average earnings by population group in South Africa. Following Deaton (1985), I used three successive cross sections to construct

[43] This also fits in with the human capital theory which predicts steeper and more concave profiles for better educated individuals (Becker, 1993).

[44] Cohort wages were also regressed on either only age or cohort control variables. Depending on the order of the polynomial, concave or humped-shaped cohort effects could be identified but never constantly declining ones. Age effects developed similarly when used as the only regressor.

a synthetic panel, with cohort means of African and White workers in formal employment replacing individual observations. The average wage per cohort was calculated to examine earnings at a disaggregated level. Preparing the data in that way enables a better utilisation of individual information provided by the October Household Surveys as well as to study changes over time.

The breakdown of overall mean wages into several age groups demonstrated that African women form the most heterogeneous group. Wages varied considerably between workers of different age but also when tracing single cohorts over time. Unlike White women, who on average experienced only small scale wage changes both across and within cohorts. In a similar way, cohort specific wage differentials were looked at. The interest was to see whether or not the gaps emerging between cohorts of different population groups remained relatively constant when moving from younger to older cohorts. Especially the comparisons by gender pointed to changing magnitudes of the differentials. Regarding Africans, cohort wages seem to diverge for older workers whereas in case of Whites the greatest discrepancies between male and female cohort wages were found among middle-aged workers.

As cohorts are followed over time it was also possible to have a closer look at within cohort changes. Although the time period of five years is rather short, different trends for younger and older cohorts could still be detected and were in most cases as expected. Young cohorts realised statistically significant wage increases, except for African women. Older cohorts, instead, either faced relatively unchanged wage levels or encountered on average a wage loss.

The limited number of periods became again crucial to the last analysis. Earnings of Africans were decomposed into age, cohort and year effects to separate life-cycle from generational effects. In the present setup, different cohorts are observed at the same age for a very short period only, causing difficulties as transitory shocks are hard to distinguish from long term trends. Especially the estimated cohort effect of African females is highly controversial as it suggests that older cohorts benefitted from generational trends. Although some features of formal sector employment in South Africa as well as the performance of the overall economy seem to be in line with this (constant or increasing real wages despite rising unemployment, low entrance rates, stagnating economy), an increase in the number of periods is compulsory to arrive at assured results.

With a time series of cross sections available, the construction of cohorts which in turn can be used to synthesise a panel structure provides a good opportunity to address temporal developments also in the absence of genuine panel data. The last decade has seen the development of consistent estimators for static as well as dynamic models. A recent study by Fitzenberger and Wunderlich (2003) tackles one of the remaining short-comings of cohort data, namely that within single cohorts an equal distribution had still to be assumed. Quantile regressions explicitly take into account the movement of the entire distribution which finally allows to use all the variation of the individual data for the estimation.

Table 1: **Cohort Definition and Cohort Size: Five-Year Age Band**

| No. of Cohorts | Age in 1995 | Men 1995 | Men 1997 | Men 1999 | Women 1995 | Women 1997 | Women 1999 |
|---|---|---|---|---|---|---|---|
| **Africans** | | | | | | | |
| 1 | 20-24 | 462 | 594 | 647 | 222 | 313 | 352 |
| 2 | 25-29 | 1029 | 989 | 931 | 508 | 570 | 503 |
| 3 | 30-34 | 1185 | 1165 | 989 | 551 | 632 | 529 |
| 4 | 35-39 | 1232 | 1050 | 799 | 565 | 583 | 472 |
| 5 | 40-44 | 939 | 762 | 588 | 377 | 408 | 286 |
| 6 | 45-49 | 807 | 591 | 392 | 322 | 292 | 206 |
| **Whites** | | | | | | | |
| 1 | 20-24 | 173 | 156 | 129 | 163 | 142 | 119 |
| 2 | 25-29 | 218 | 172 | 154 | 172 | 131 | 129 |
| 3 | 30-34 | 278 | 203 | 165 | 174 | 118 | 95 |
| 4 | 35-39 | 289 | 183 | 127 | 171 | 120 | 91 |
| 5 | 40-44 | 248 | 144 | 103 | 139 | 109 | 88 |
| 6 | 45-49 | 217 | 157 | 88 | 124 | 89 | 67 |

Table 2: **Cohort Definition and Cohort Size: Two-Year Age Band**

| No. of Cohorts | Age in 1995 | Men 1995 | Men 1997 | Men 1999 | Women 1995 | Women 1997 | Women 1999 |
|---|---|---|---|---|---|---|---|
| **Africans** | | | | | | | |
| 1 | 21-22 | 162 | 232 | 233 | 70 | 108 | 145 |
| 2 | 23-24 | 257 | 273 | 325 | 120 | 165 | 140 |
| 3 | 25-26 | 356 | 363 | 389 | 194 | 215 | 223 |
| 4 | 27-28 | 457 | 419 | 349 | 208 | 250 | 198 |
| 5 | 29-30 | 502 | 456 | 388 | 242 | 247 | 177 |
| 6 | 31-32 | 453 | 481 | 399 | 222 | 236 | 219 |
| 7 | 33-34 | 446 | 435 | 395 | 193 | 254 | 215 |
| 8 | 35-36 | 558 | 451 | 359 | 255 | 241 | 204 |
| 9 | 37-38 | 465 | 438 | 309 | 232 | 238 | 197 |
| 10 | 39-40 | 507 | 342 | 236 | 171 | 195 | 134 |
| 11 | 41-42 | 338 | 306 | 257 | 147 | 158 | 137 |
| 12 | 43-44 | 301 | 275 | 226 | 137 | 159 | 86 |
| 13 | 45-46 | 415 | 279 | 193 | 166 | 139 | 115 |
| 14 | 47-48 | 260 | 226 | 134 | 110 | 111 | 64 |

Table 3: **Real Hourly Mean Wage by Cohort, 1995 and 1999**

| Age in | | Wage of Cohort in | |
|---|---|---|---|
| 1995 | 1999 | 1995 | 1999 |
| **African Men** | | | |
| 24-27 | 28-31 | 7.32 | 8.25 |
| 28-31 | 32-35 | 8.50 | 8.19 |
| 32-35 | 36-39 | 9.02 | 10.66 |
| 36-39 | 40-43 | 9.23 | 10.47 |
| 40-43 | 44-47 | 9.91 | 9.40 |
| 44-47 | 48-51 | 9.14 | 8.93 |
| 48-51 | 52-55 | 9.53 | 8.13 |
| 52-55 | 56-59 | 8.94 | 7.91 |
| **African Women** | | | |
| 24-27 | 28-31 | 7.98 | 8.58 |
| 28-31 | 32-35 | 9.05 | 9.52 |
| 32-35 | 36-39 | 10.14 | 10.06 |
| 36-39 | 40-43 | 10.11 | 11.30 |
| 40-43 | 44-47 | 9.59 | 8.60 |
| 44-47 | 48-51 | 9.28 | 8.99 |
| 48-51 | 52-55 | 8.58 | 7.82 |
| 52-55 | 56-59 | 7.99 | 7.54 |
| **White Men** | | | |
| 24-27 | 28-31 | 20.81 | 27.02 |
| 28-31 | 32-35 | 26.23 | 33.14 |
| 32-35 | 36-39 | 31.83 | 34.32 |
| 36-39 | 40-43 | 32.98 | 32.39 |
| 40-43 | 44-47 | 34.32 | 32.56 |
| 44-47 | 48-51 | 41.67 | 33.05 |
| 48-51 | 52-55 | 35.77 | 35.24 |
| 52-55 | 56-59 | 37.35 | 35.87 |
| **White Women** | | | |
| 24-27 | 28-31 | 15.39 | 18.88 |
| 28-31 | 32-35 | 18.04 | 24.87 |
| 32-35 | 36-39 | 16.11 | 20.54 |
| 36-39 | 40-43 | 19.99 | 19.85 |
| 40-43 | 44-47 | 21.24 | 22.62 |
| 44-47 | 48-51 | 20.57 | 23.50 |
| 48-51 | 52-55 | 20.06 | 20.10 |
| 52-55 | 56-59 | 17.98 | 23.27 |

Numbers shown are sample weighted means (1995 prices).

Table 4: **Decomposition Analysis for Africans**

| | African Men | | | | | | African Women | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **(1)** | | **(2)** | | **(3)** | | **(1)** | | **(2)** | | **(3)** | |
| Intercept | 7.63** | (1.37) | -8.16 | (6.07) | -0.88 | (0.57) | 10.47** | (1.52) | -12.14 | (6.41) | 0.50 | (0.68) |
| Education | - | | 1.95* | (0.74) | - | | - | | 2.48** | (0.69) | - | |
| *Age polynomial* | | | | | | | | | | | | |
| Age | 0.56** | (0.08) | 0.42** | (0.09) | 0.47** | (0.06) | 0.26** | (0.08) | 0.17* | (0.07) | 0.26** | (0.07) |
| $Age^2$ | -0.09** | (0.02) | -0.07** | (0.02) | -0.04** | (0.01) | -0.03** | (0.01) | -0.02* | (0.01) | -0.02** | (0.01) |
| $Age^3$ | 0.61** | (0.14) | 0.54** | (0.14) | 0.14** | (0.03) | 0.09 | (0.05) | 0.06 | (0.04) | 0.05** | (0.02) |
| $Age^4$ | -0.02** | (0.00) | -0.01** | (0.00) | 0.00** | (0.00) | - | | - | | - | |
| *Cohort polynomial* | | | | | | | | | | | | |
| Cohort | 0.05 | (0.06) | 0.14* | (0.06) | -0.02 | (0.06) | - | | - | | - | |
| $Cohort^2$ | -0.04 | (0.86) | -0.83 | (0.84) | 0.14 | (0.49) | - | | - | | - | |
| $Cohort^3$ | -0.01 | (0.04) | 0.01 | (0.04) | 0.00 | (0.01) | - | | - | | - | |
| *Cohort dummy variables* | | | | | | | | | | | | |
| 17-18 | - | | - | | - | | - | | - | | 0.27 | (0.20) |
| 19-20 | - | | - | | - | | - | | - | | 0.48* | (0.21) |
| 21-22 | - | | - | | - | | - | | - | | 0.63** | (0.23) |
| 23-24 | - | | - | | - | | 0.20 | (0.11) | 0.16 | (0.09) | 0.75** | (0.24) |
| 25-26 | - | | - | | - | | 0.43** | (0.13) | 0.44** | (0.11) | 0.89** | (0.26) |
| 27-28 | - | | - | | - | | 0.46** | (0.15) | 0.51** | (0.12) | 0.93** | (0.27) |
| 29-30 | - | | - | | - | | 0.58** | (0.16) | 0.63** | (0.13) | 1.06** | (0.28) |
| 31-32 | - | | - | | - | | 0.64** | (0.18) | 0.70** | (0.14) | 1.11** | (0.29) |
| 33-34 | - | | - | | - | | 0.81** | (0.19) | 0.83** | (0.15) | 1.27** | (0.29) |
| 35-36 | - | | - | | - | | 0.98** | (0.20) | 1.00** | (0.16) | 1.39** | (0.30) |
| 37-38 | - | | - | | - | | 1.06** | (0.21) | 1.08** | (0.17) | 1.45** | (0.30) |
| 39-40 | - | | - | | - | | 1.21** | (0.22) | 1.23** | (0.18) | 1.64** | (0.30) |
| 41-42 | - | | - | | - | | 1.29** | (0.24) | 1.27** | (0.19) | 1.72** | (0.30) |
| 43-44 | - | | - | | - | | 1.32** | (0.25) | 1.32** | (0.20) | 1.76** | (0.31) |
| 45-46 | - | | - | | - | | 1.54** | (0.27) | 1.52** | (0.22) | 1.86** | (0.31) |
| 47-48 | - | | - | | - | | 1.55** | (0.29) | 1.52** | (0.23) | 2.00** | (0.32) |

*continued on next page*

Table 4: *continued*

| | African Men | | | | | | African Women | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **(1)** | | **(2)** | | **(3)** | | **(1)** | | **(2)** | | **(3)** | |
| 49-50 | - | | - | | - | | - | | - | | $2.15^{**}$ | (0.33) |
| 51-52 | - | | - | | - | | - | | - | | $2.24^{**}$ | (0.34) |
| 53-54 | - | | - | | - | | - | | - | | $2.29^{**}$ | (0.35) |
| 55-56 | - | | - | | - | | - | | - | | $2.33^{**}$ | (0.38) |
| 57-58 | - | | - | | - | | - | | - | | $2.54^{**}$ | (0.41) |
| 59-60 | - | | - | | - | | - | | - | | $2.59^{**}$ | (0.45) |
| *Year* | | | | | | | | | | | | |
| 1999 | $-0.02^{*}$ | (0.01) | -0.01 | (0.01) | $-0.02^{**}$ | (0.01) | 0.01 | (0.01) | 0.02 | (0.01) | $0.02^{*}$ | (0.01) |
| N | 42 | | 42 | | 69 | | 42 | | 42 | | 69 | |
| $R^2$ | 0.97 | | 0.98 | | 0.99 | | 0.95 | | 0.97 | | 0.99 | |

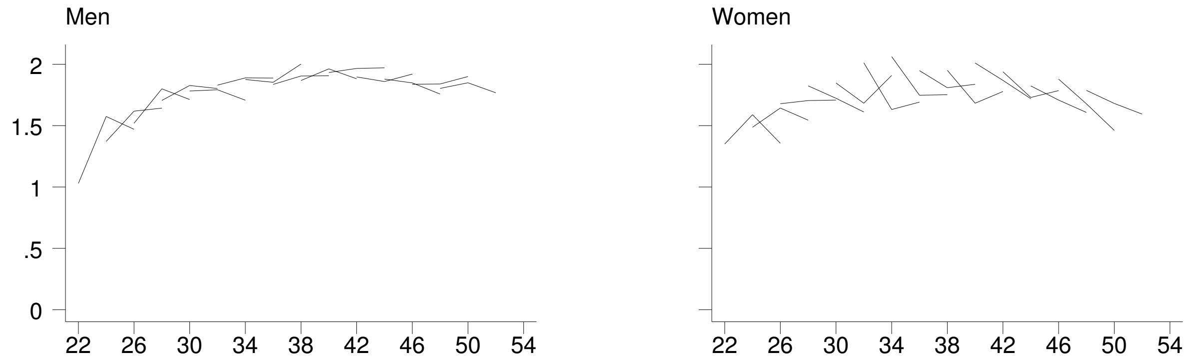
Significance levels: $*$ : 5% $**$ : 1%. Standard errors in parentheses.

Cohort dummy variables correspond to the age in 1995. Reference categories: Specifications (1), (2): Cohort 20-21, Specification (3): Cohort 15-16.

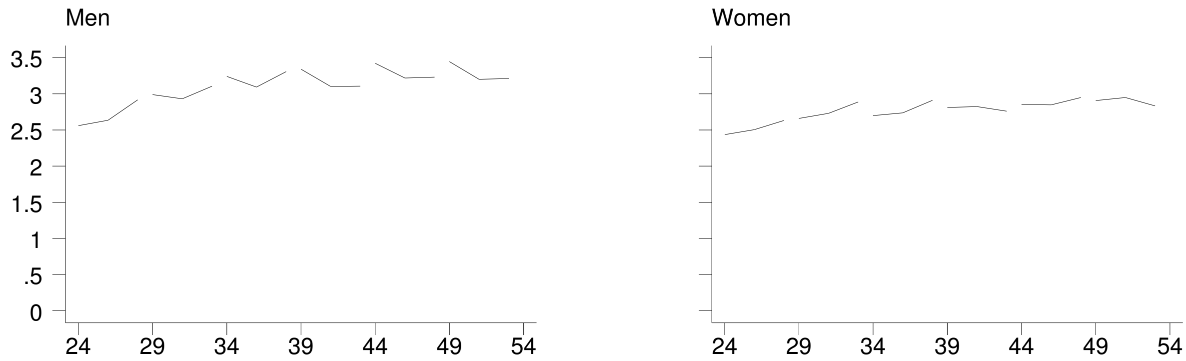
Education is measured in years of schooling completed.

Variables $\text{Age}^3$, $\text{Age}^4$, $\text{Cohort}^2$, and $\text{Cohort}^3$ have been scaled by $10^{-2}$.

Figure 1: **Logarithm of Real Hourly Wage (1995 prices) by Cohort**



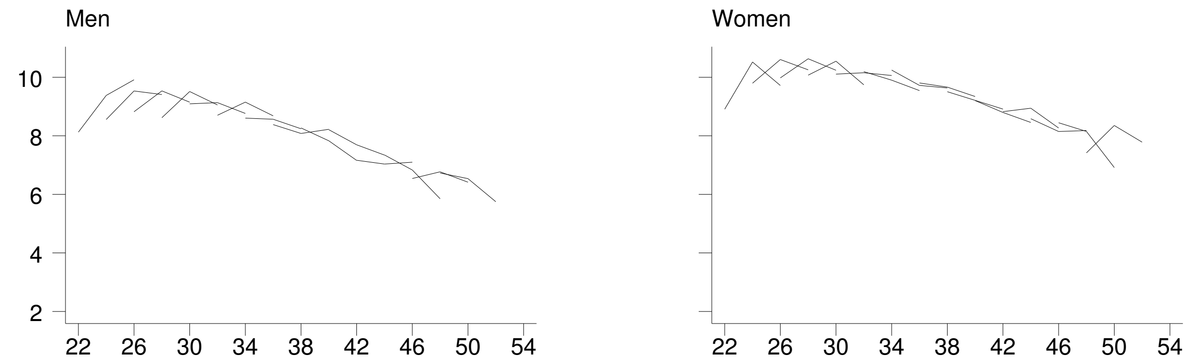

(a) African workers.



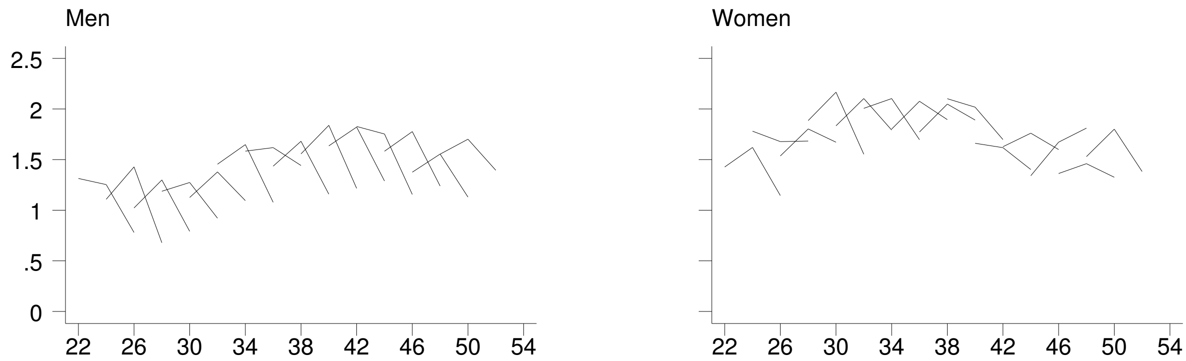

(b) White workers.

*Notes:* Figures are based on sample weighted cohort means. The x-axis in each graph is labelled according to the upper age limit of the individual cohorts.

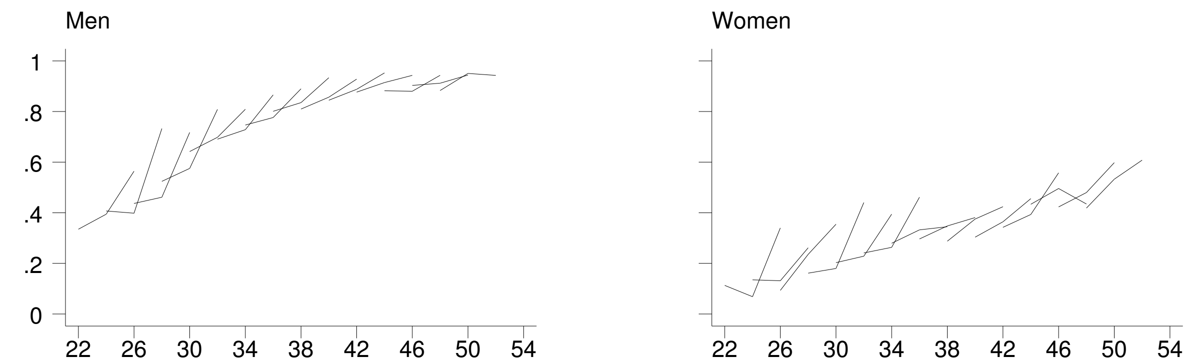
Figure 2: **Summary Statistics by Cohort, African Workers**




(a) Number of years of schooling completed.



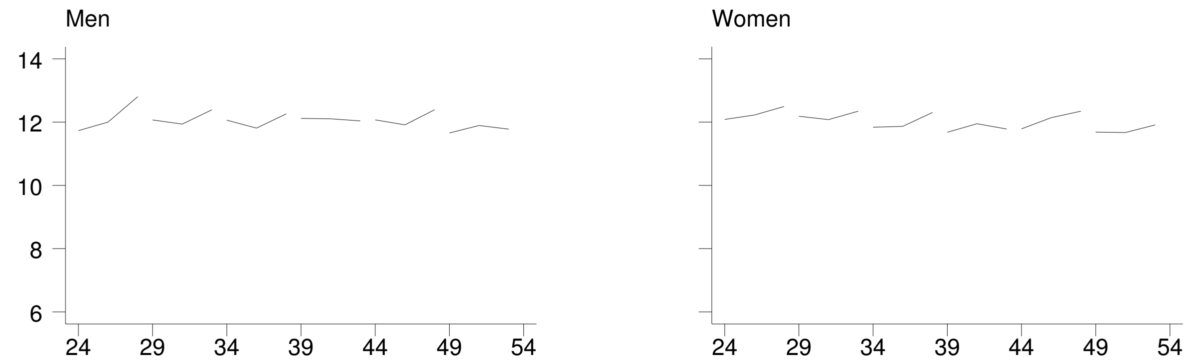

(b) Number of children living in the worker's household.



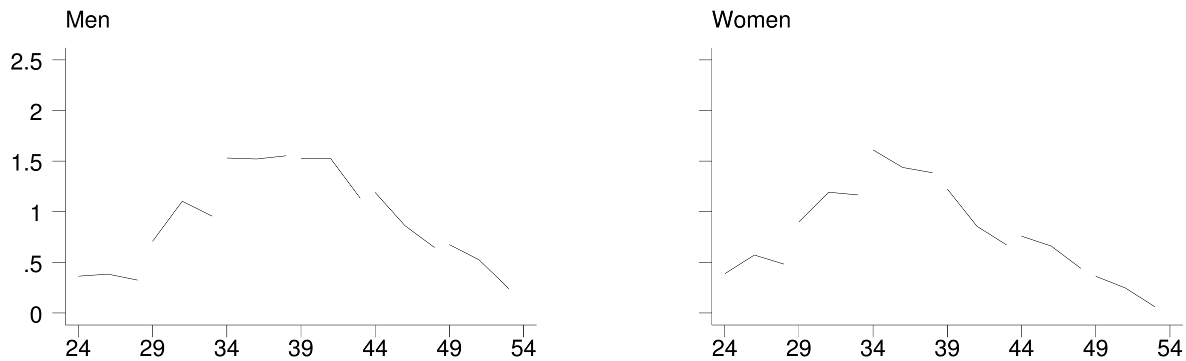

(c) Proportion of workers acting as household head.

*Notes:* See Figure 1.

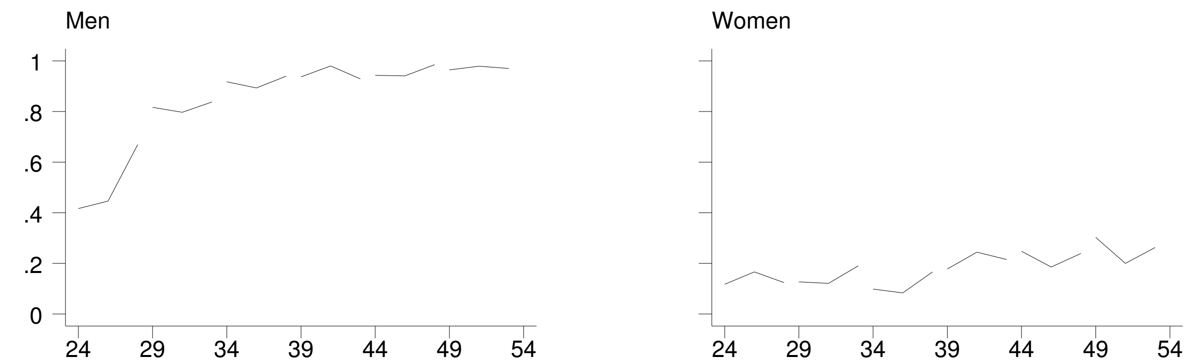
Figure 3: **Summary Statistics by Cohort, White Workers**




(a) Number of years of schooling completed.



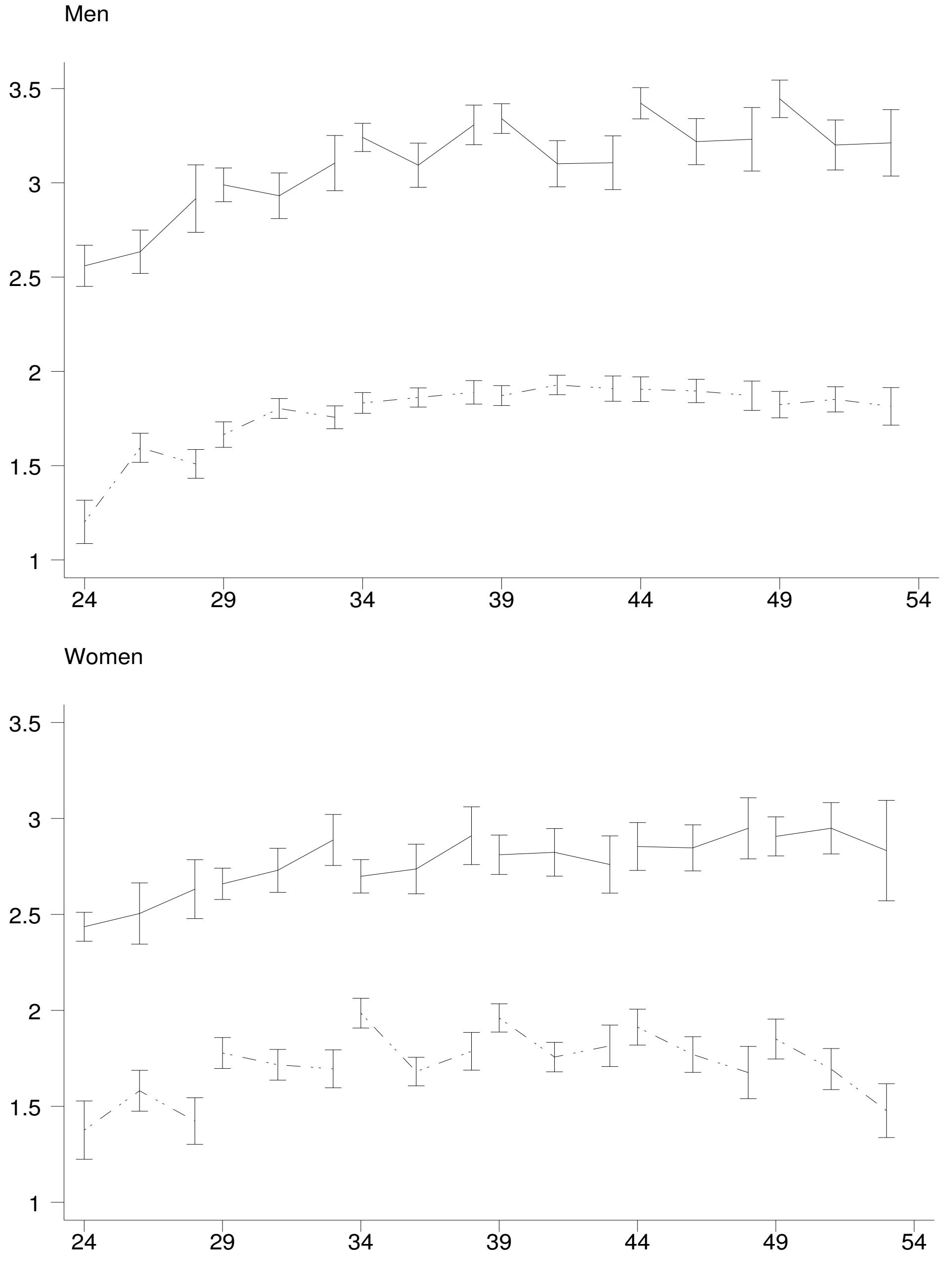

(b) Number of children living in the worker's household.




(c) Proportion of workers acting as household head.

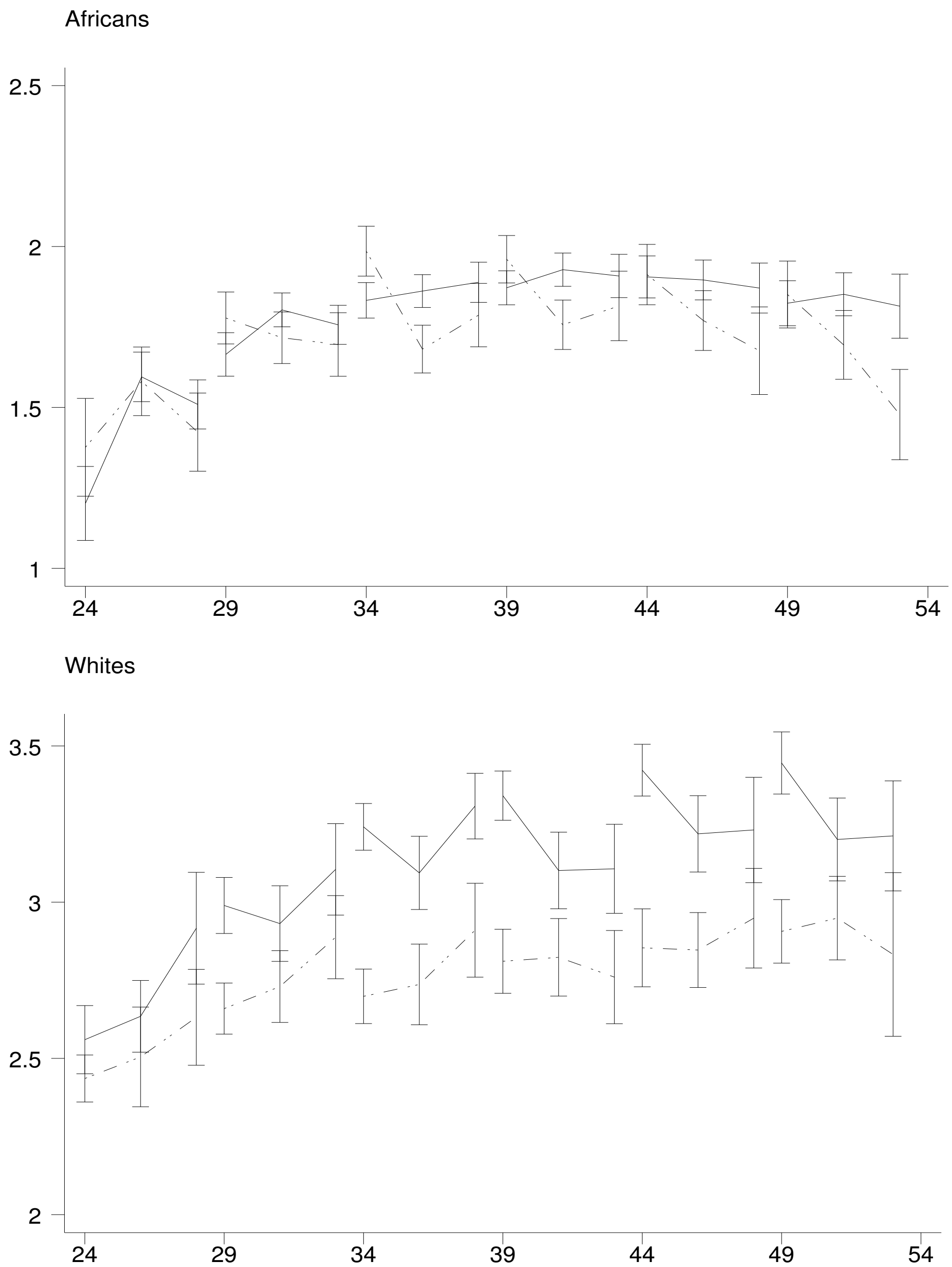
*Notes:* See Figure 1.

Figure 4: **Racial Wage Differentials by Cohort**




*Notes:* See Figure 1. Solid lines correspond to cohort wages of White workers, dashed lines to those of African workers. Cohort wages are enclosed by confidence bands of 95 per cent.

Figure 5: **Gender Wage Differentials by Cohort**




*Notes:* See Figure 1. Solid lines correspond to cohort wages of male workers, dashed lines to those of female workers. Cohort wages are enclosed by confidence bands of 95 per cent.

Figure 6: **Age, Cohort, and Year Effects of Wages Earned by African Men (1)**

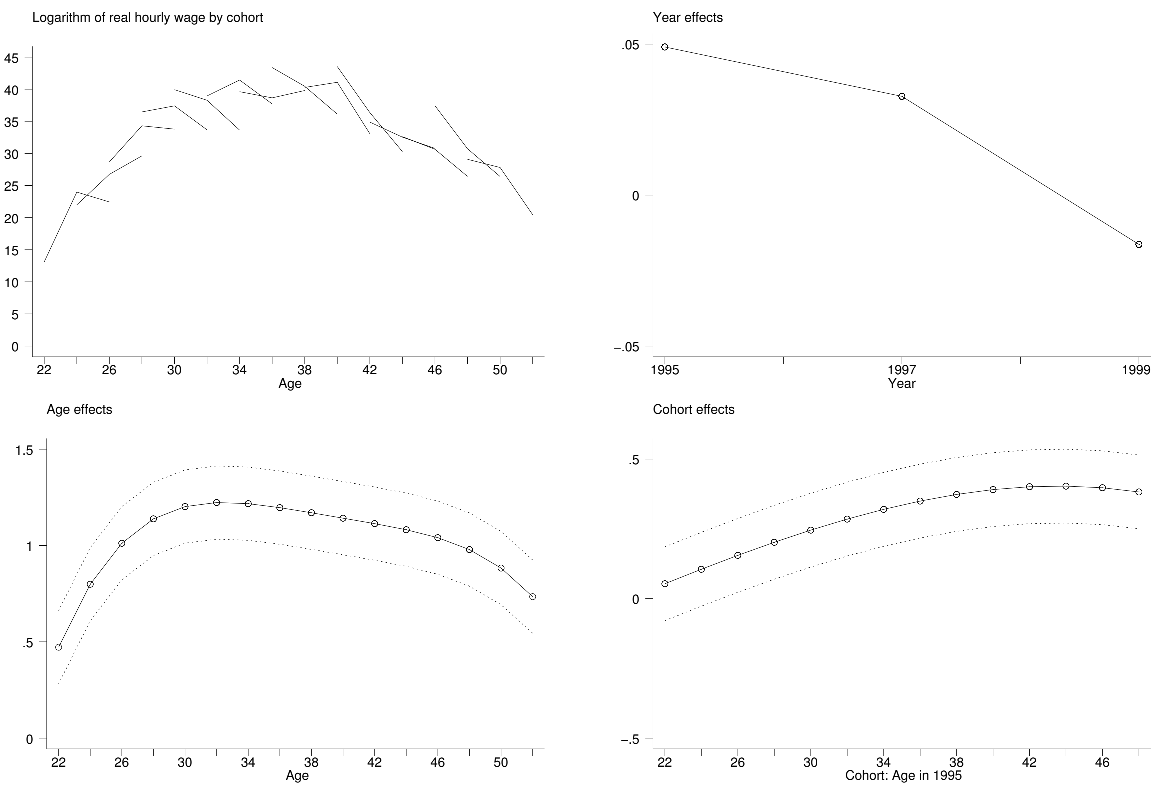


Figure 7: **Age, Cohort, and Year Effects of Wages Earned by African Women (1)**

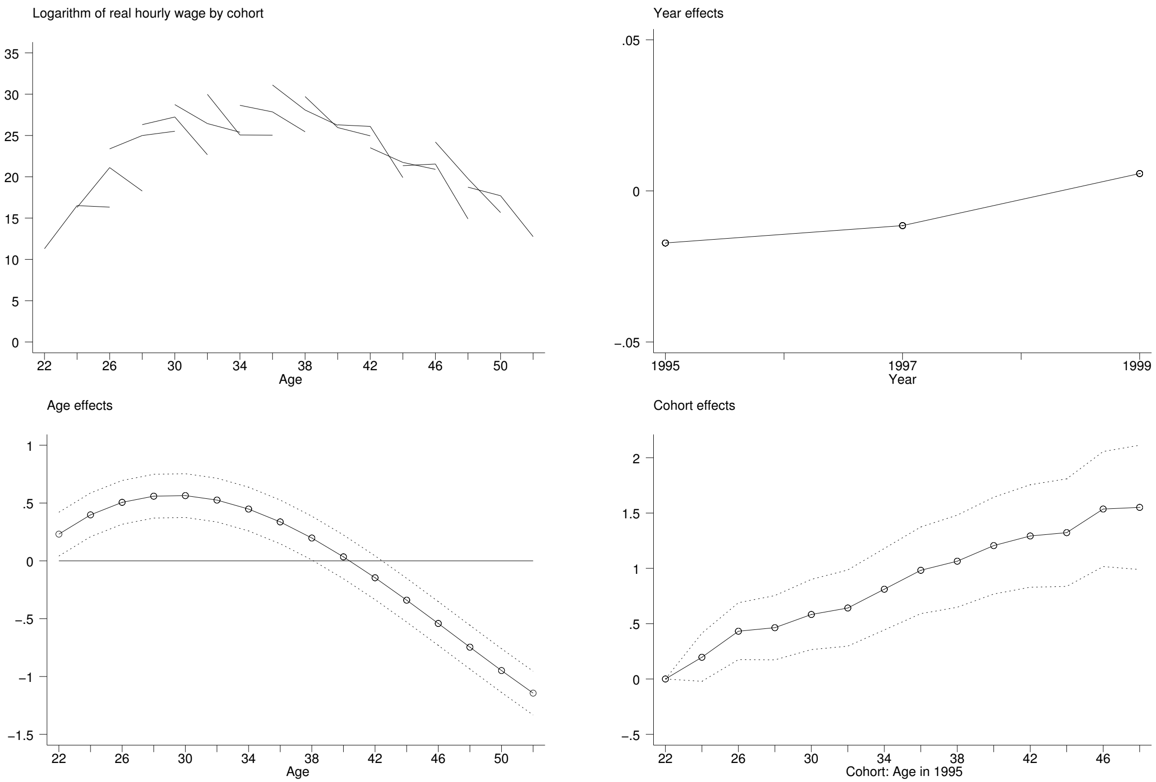


Figure 8: **Age, Cohort, and Year Effects of Wages Earned by African Men (2)**

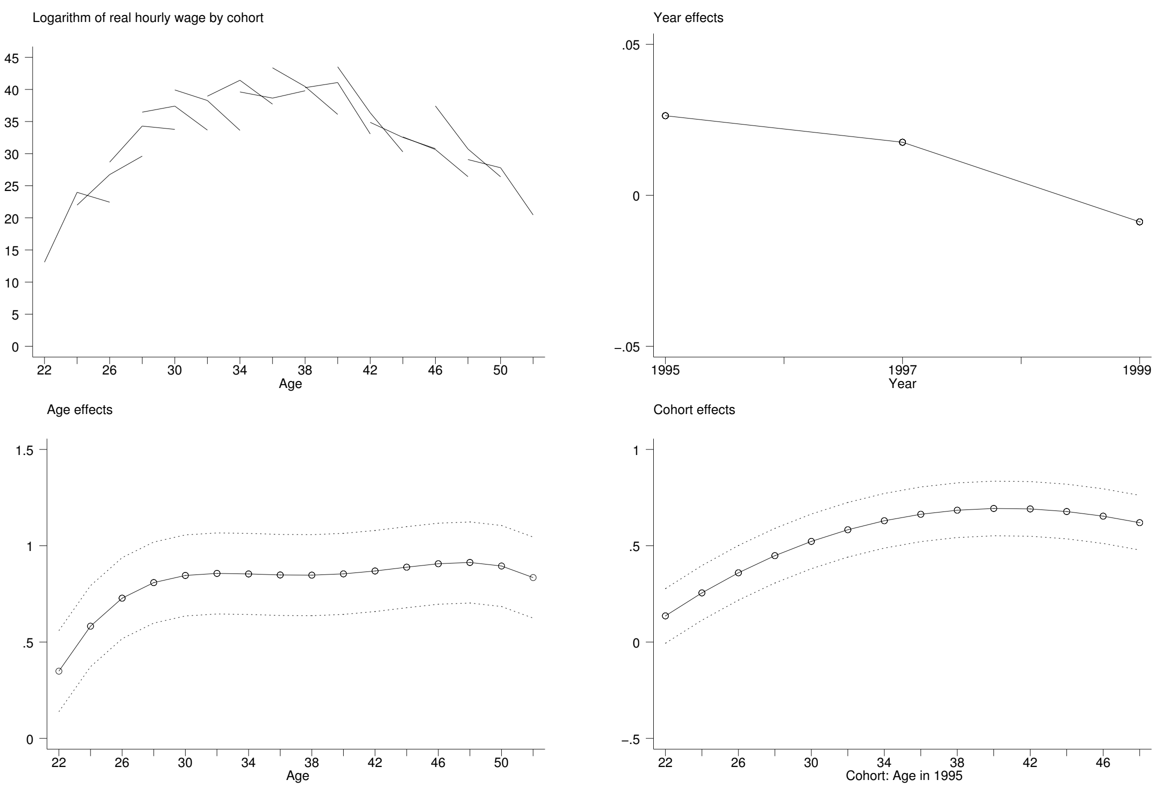


Figure 9: **Age, Cohort, and Year Effects of Wages Earned by African Women (2)**

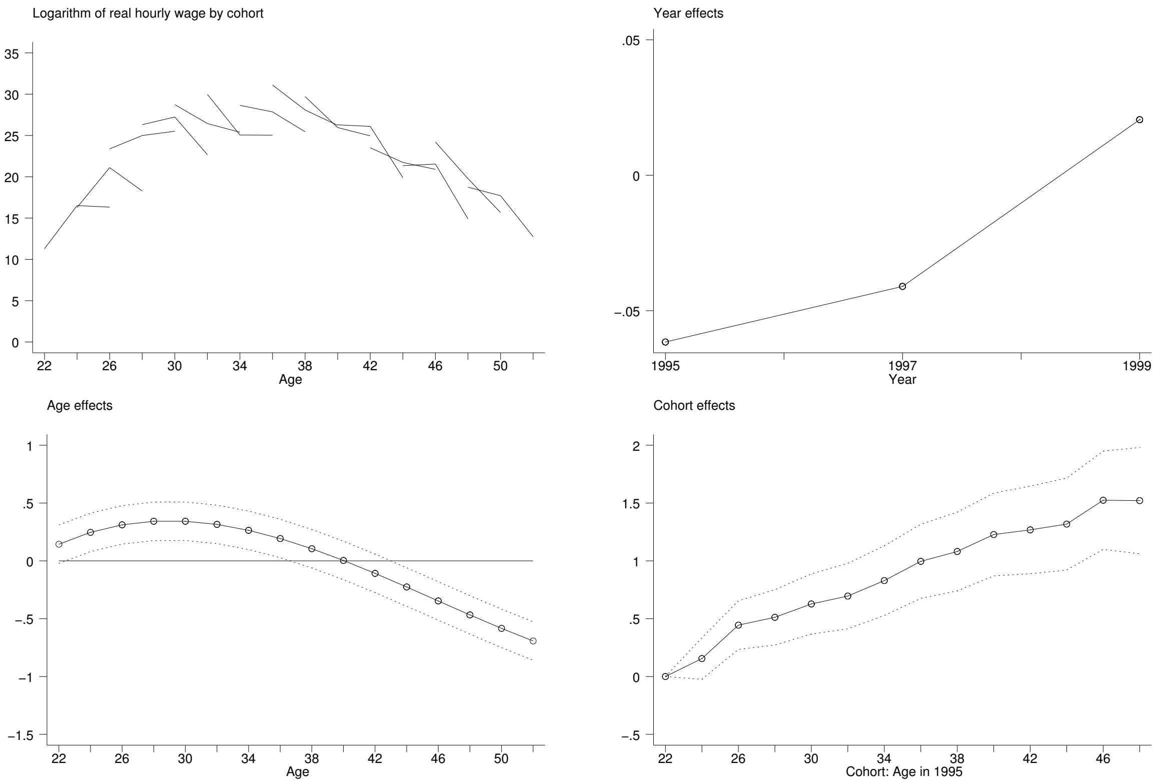


Figure 10: **Age, Cohort, and Year Effects of Wages Earned by African Men (3)**

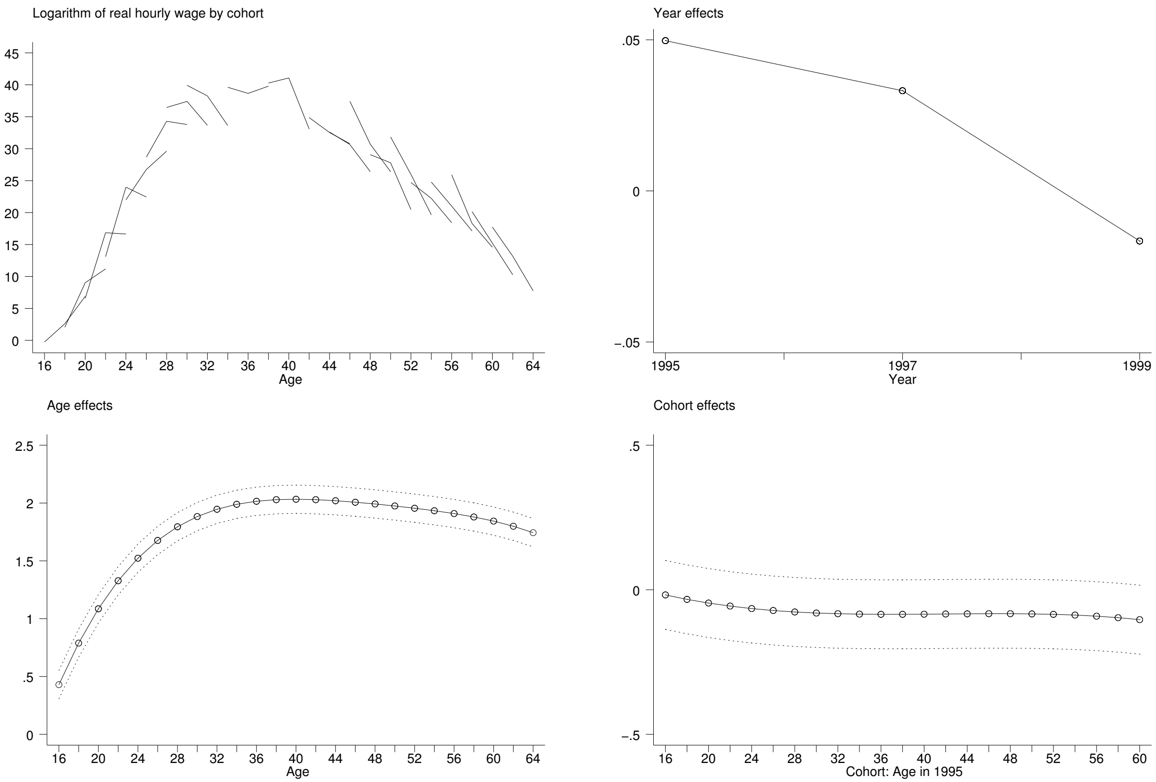
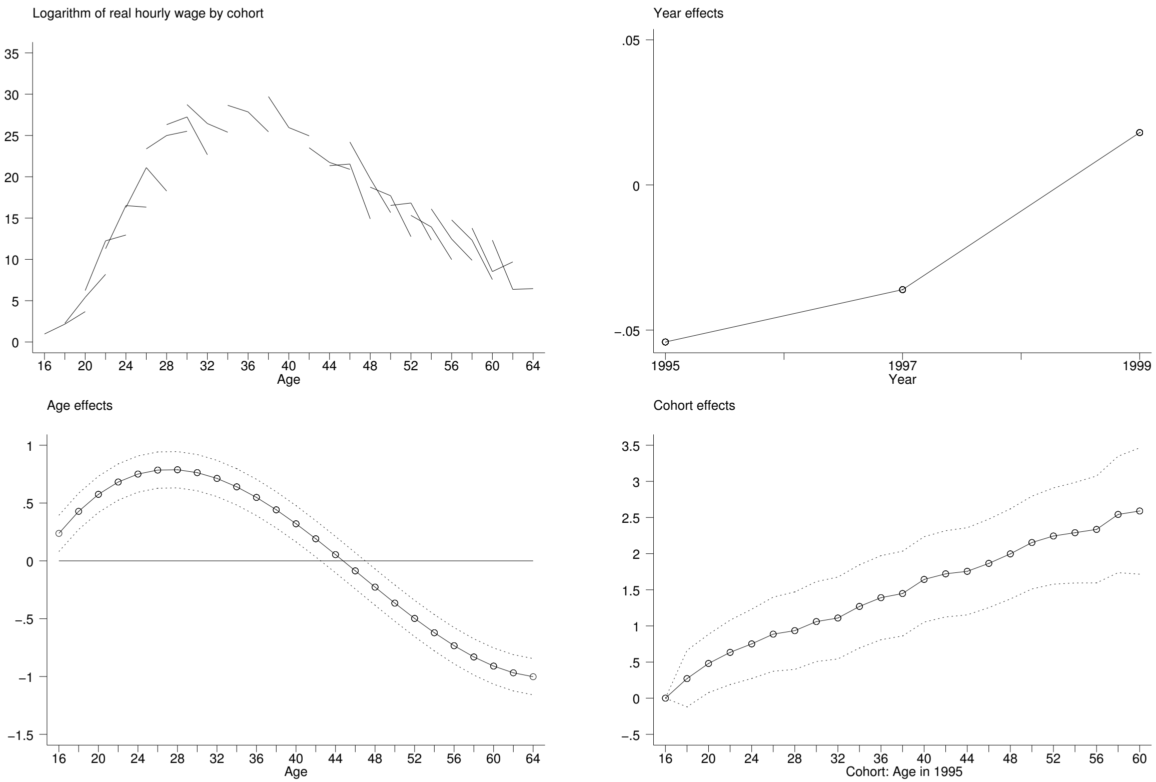


Figure 11: **Age, Cohort, and Year Effects of Wages Earned by African Women (3)**



# References


Allanson, P., and J. P. Atkins (2001): "Labour market reform and the evolution of the racial wage hierarchy in Post-Apartheid South Africa," DPRU Working Paper No. 01/59, University of Cape Town.

Allanson, P., J. P. Atkins, and T. Hinks (2000a): "Did the end of Apartheid spell the beginning of the end for the racial wage hierarchy in South Africa? A multilateral analysis of racial wage differentials in the early Post-Apartheid period," Department of Economic Studies Discussion Paper No. 118, University of Dundee.

——— (2000b): "A multilateral decomposition of racial wage differentials in the 1994 South African labour market," *Journal of Development Studies*, 37(1), 93–120.

Attanasio, O. (1997): "Cohort analysis of saving behavior by U.S. households," *Journal of Human Resources*, 33(3), 575–609.

Attanasio, O., G. Berloffa, R. Blundell, and I. Preston (2002): "From earnings inequality to consumption inequality," *Economic Journal*, 112(478), C52–C59.

Attanasio, O., and G. Weber (1994): "The U.K. consumption boom of the late 1980s: Aggregate implications of microeconomic evidence," *Economic Journal*, 104(427), 1269–1302.

Baltagi, B. H. (1995): *Econometric analysis of panel data*. John Wiley, New York.

——— (1998): "Panel Data Methods," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah, and D. Giles, vol. 155 of *Statistics: Textbooks and Monographs*, pp. 291–323. Marcel Dekker, New York, Basel and Hong Kong.

Bardazzi, R. (2000): "Effects of the changing structure of population on the Italian economy," INFORUM Working Papers No. 00-005, University of Maryland.

Becker, G. S. (1993): *Human Capital*. University of Chicago Press, Chicago, third edn.

Blau, F. D., M. A. Ferber, and A. E. Winkler (1998): *The Economics of Women, Men, and Work*. Prentice Hall International (UK) Limited, London, third edn.

Blundell, R., M. Browning, and C. Meghir (1994): "Consumer demand and the life-cycle allocation of household expenditures," *Review of Economic Studies*, 61(1), 57–80.

Blundell, R., C. Meghir, and P. Neves (1993): "Labour supply and intertemporal substitution," *Journal of Econometrics*, 59, 137–160.

Browning, M., A. Deaton, and M. Irish (1985): "A profitable approach to labor supply and commodity demands over the life-cycle," *Econometrica*, 53(3), 503–543.

Cichello, P. L., G. S. Fields, and M. Leibbrandt (2002): "Earnings and employment dynamics for Africans in Post-Apartheid KwaZulu-Natal," Paper prepared for DPRU (FES) second annual conference on labour markets and poverty in South Africa, Johannesburg.

Collado, M. D. (1997): "Estimating dynamic models from time series of cross-sections," *Journal of Econometrics*, 82(1), 37–62.

Deaton, A. (1985): "Panel Data from time series of cross-sections," *Journal of Econometrics*, 30, 109–126.

——— (1997): *The Analysis of Household Surveys*. The John Hopkins University Press, Baltimore.

Deaton, A., and C. Paxson (1994a): "Intertemporal choice and inequality," *Journal of Political Economy*, 102(3), 437–467.

——— (1994b): "Saving, growth and aging in Taiwan," in *Studies in the Economics of Aging*, ed. by D. A. Wise. NBER Project Report, University of Chicago Press, Chicago.

Erichsen, G., and J. Wakeford (2001): "Racial wage discrimination in South Africa before and after the first democratic election," DPRU Working Paper No. 01/49, University of Cape Town.

Fallon, P., and R. Lucas (1998): "South African labor markets adjustment and inequalities," Discussion Paper No. 12, The World Bank Southern Africa Department, Washington, DC.

Fedderke, J., and M. Mariotti (2002): "Changing labour market conditions in South Africa: A sectoral analysis of the period 1970-1997," *The South African Journal of Economics*, 70(5).

Fitzenberger, B., and G. Wunderlich (2003): "Gender wage differences in West Germany: A cohort analysis," *German Economic Review*, 3(4), 379–414.

Gohmann, S. F., M. J. McCrickard, and F. Slesnick (1998): "Age-earnings profiles estimates: Do they change over time?," *Journal of Forensic Economics*, 11(3), 173–188.

Greene, W. H. (1997): *Econometric Analysis*. Prentice Hall International Limited, London, third edn.

——— (2000): *Econometric Analysis*. Prentice Hall International Limited, London, fourth edn.

Grün, C. (2003): "Direct and Indirect Labour Market Discrimination in the South African Labour Market," Mimeographed, University of Munich.

Heckman, J. J., and R. Robb (1985): "Using longitudinal data to estimate age, period and cohort effects in earnings equation," in *Cohort Analysis in Social Research - Beyond the Identification Problem*, ed. by W. M. Mason, and S. E. Fienberg, pp. 137–150. Springer Verlag, New York.

Hinks, T. (2002): "Gender wage differentials and discrimination in the new South Africa," *Applied Economics*, 34(16), 2043–2052.

Jappelli, T. (1999): "The age-wealth profile and the life-cycle hypothesis: A cohort analysis with a time series of cross-sections of Italian households," *Review of Income and Wealth*, 45(1), 57–75.

Jappelli, T., and F. Modigliani (1998): “The aging-saving profile and the life-cycle hypothesis,” CSEF Working Paper No. 9, Centre for Studies in Economics and Finance, Salerno.

Jenkins, S. P. (1994): “Earnings discrimination measurement: A distributional approach,” *Journal of Econometrics*, 61(1), 81–102.

Johnson, G. E., and F. P. Stafford (1974): “Lifetime earnings in a professional labor market: Academic economists,” *Journal of Political Economy*, 82(3), 549–569.

Kennedy, P. (1998): *A Guide to Econometrics*. MIT Press, Cambridge, MA, fourth edn.

Keswell, M. (2000): “Labour market dynamics in South Africa: Evidence from KwaZulu-Natal Province,” Mimeographed, University of Massachusetts.

Kingdon, G., and J. Knight (2001): “Race and incidence of unemployment in South Africa,” CSAE Working Paper No. WPS/2001.18, Centre for the Study of African Economies, Oxford.

Kitamura, Y. (2001): “An econometric analysis of cohort data from household savings in Japan,” Mimeographed, Hitotsubashi University, Tokyo.

Klasen, S., and I. Woolard (2000): “Surviving unemployment without state support: Unemployment and household formation in South Africa,” SFB 386 Discussion Paper No. 213, University of Munich.

——— (2002): “Income mobility and household dynamics in South Africa,” Mimeographed, University of Munich.

May, J., M. R. Carter, L. Haddad, and J. Maluccio (2000): “KwaZulu-Natal Income Dynamics Study (KIDS) 1993-1998: A longitudinal household database for South African policy analysis,” *Development Southern Africa*, 17(4), 567–581.

Mincer, J. (1962): “On-the-job training: Costs, returns and some implications,” *Journal of Political Economy*, 70(5), S50–S79.

Modigliani, F. (1986): “Life cycle, individual thrift, and the wealth of nations,” *American Economic Review*, 76(3), 297–313.

Moffitt, R. (1993): “Identification and estimation of dynamic models with a time series of cross-sections,” *Journal of Econometrics*, 59(1-2), 99–123.

Ohtake, F., and M. Saito (1998): “Population aging and consumption inequality in Japan,” *Review of Income and Wealth*, 44(3), 361–381.

Rodgers, J. D., M. L. Brookshire, and R. J. Thornton (1996): “Forecasting earnings using age-earnings profiles and longitudinal data,” *Journal of Forensic Economics*, 9(2), 169–210.

Rospabe, S. (2001a): “An empirical evaluation of gender discrimination in employment, occupation attainment and wage in South Africa in the late 1990s,” Mimeographed, University of Cape Town.

——— (2001b): "Making racial wage relations fair in South Africa: A focus on the role of trade unions," DPRU Working Paper No. 01/48, University of Cape Town.

——— (2002): "How did labour market racial discrimination evolve after the end of Apartheid?," *The South African Journal of Economics*, 70(1), 185–217.

Shorrocks, A. F. (1975): "The age-wealth relationship: A cross-section and cohort analysis," *Review of Economics and Statistics*, 57(2), 155–163.

Stats SA (2001): "South African Statistics 2001," http://www.statssa.gov.za, Statistics South Africa, Pretoria.

Verbeek, M. (1996): "Pseudo Panel Data," in *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, vol. 33 of *Advanced Studies in Theoretical and Applied Econometrics*, pp. 280–292. Kluwer Academic, Dordrecht, Boston and London, 2nd edn.

Verbeek, M., and T. Nijman (1992): "Can cohort data be treated as genuine panel data?," *Empirical Economics*, 17(1), 9–23.

——— (1993): "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections," *Journal of Econometrics*, 59(1-2), 125–136.

Verbeek, M., and F. Vella (2000): "Estimating dynamic models from repeated cross-sections," Discussion Paper No. 2000-25, Center for Economic Research, Tilburg University.

WDI (2002): "World Development Indicators 2002," The World Bank, Washington, DC.