

CSAE Working Paper WPS/2012-04

The Determinants of Earnings Inequalities: Panel data evidence from South Africa

Andrew Kerr* and Francis Teal†

February 29, 2012

Abstract

In this paper we analyse the relative importance of individual ability and labour market institutions, including public sector wage setting and trade unions, in determining earnings differences across different types of employment. To do this we use the KwaZulu-Natal Income Dynamics Study data from South Africa, which show extremely large average earnings differentials across different types of employment. Our results suggest that human capital and individual ability explain much of the earnings differentials within the private sector, including the union premium, but cannot explain the large premiums for public sector workers. We show that a public sector premium exists only for those moving into the public sector. The paper addresses the challenges of non-random attrition and measurement error bias that panel data bring. Our results show that emphasising a simple binary dichotomy between the formal and informal sector can be unhelpful in attempting to explore how the labour market functions.

JEL Classification: J31, J51, J45, O12

Keywords: Formality, Trade unions, Public sector, Earnings, South Africa.

* Centre for the Study of African Economies and Department of Economics, University of Oxford and DataFirst, University of Cape Town, Private Bag X3, Rondebosch, 7701, Cape Town, South Africa (Present Address). andrew.kerr@uct.ac.za (corresponding author). Telephone: +27216504952, Fax: +27216505709

† Centre for the Study of African Economies and Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom. francis.teal@economics.ox.ac.uk

1 Introduction

In 2004 black self-employed workers in KwaZulu-Natal province, South Africa, had average hourly earnings six times lower than black employees in private sector firms, who themselves earned half the average hourly wage of black workers in the public sector. In this paper we seek to explain the vast differences between the average earnings of the self-employed, at one end of the informal/formal divide in South Africa, and unionised public sector workers at the other.

Human capital theory provides one potential explanation for earnings differentials, positing that workers' productivities explain their earnings, and sector choices, and implying that sectoral choices are utility maximising, despite the possible existence of sectoral earnings differentials. Extensions and modifications of the Harris and Todaro (1970) model provide an alternative explanation, suggesting that institutional features of the labour market prevent formal sector earnings from equalising the demand and supply of labour in this sector and can generate wage differentials between different sectors for otherwise identical workers.

In this paper we test these alternative explanations using a panel data set that allows us to explore the role of observed and unobserved individual heterogeneity in determining the earnings outcomes observed in the South African labour market. The data set also allows us to analyse the relative contributions of human capital and two key formal sector institutions, public sector pay policy and trade unions, to earnings determination. In exploring the impact of these institutions, and including all income-earning individuals we go beyond the work of Badaoui et al. (2008), the only other paper that uses panel data to explore this topic in South Africa, who focus solely on wage employees in the private sector. In addition, the work of Badaoui et al. (2008) makes use of the benefits of panel data but does not address its limitations, particularly measurement error and attrition, and does not consider the endogeneity of movement between sectors. Our work addresses the possible effects of attrition and measurement error and we explore the impact of endogenous movement on our results.

As a result of the legacy of Apartheid racial classifications continue to be commonly used in surveys, censuses and in analysis of the South African labour market. We abstract from racial differentials in labour market outcomes in this paper by confining our analysis to black South Africans only. This decision mainly reflects data constraints: the panel data set we use only contains information on black and Indian households. But given the large differences in labour market outcomes in South Africa across all racial groups, we have chosen to further limit our analysis to black workers.

The structure of the paper is as follows: Section 2 reviews explanations for, and evidence on, possible explanations of earnings differentials. The KwaZulu-Natal Income Dynamics Study is described in Section 3. Section 4 tests the ability of both human capital theory and segmentation hypotheses to explain earnings, as well as analysing the effects of allowing for unobserved heterogeneity to influence earnings. Section 5 concludes.

2 Identifying Sources of Wage Differentials

2.1 Explaining Sectoral wage differentials

One explanation for the large sectoral earnings differentials in South Africa noted above posits the existence of institutional features of the labour market which prevent earnings in the formal sector from equalising the demand and supply of labour in this sector, and which generate sectoral wage differentials for otherwise identical workers, often described as segmentation. In the developing country context the Fields (1975) extension of the Harris and Todaro (1970) model of migration has been the basis for much of the empirical literature seeking to explore whether segmentation exists in the labour market. This model included an urban, informal, free-entry sector, along with the urban formal sector and rural agriculture, where the existence of a minimum wage or union activity in the formal sector created wage differentials between the formal and informal sectors and left those in the informal sector worse off than those in the formal sector.

An alternative explanation for earnings differentials is the traditional neo-classical framework, which emphasises the productivity of individuals as the primary driver of wages in a competitive labour market. In this explanation of labour market outcomes earnings differentials across different types of employment simply reflect average differences in individual human capital and ability in these different types of employment (Heckman and Sedlacek (1985), Heckman and Hotz (1986), Maloney (2004)). These explanations are inspired by the Roy (1951) model of the labour market, in which individuals' comparative advantages in either of the two sectors determines their choice of where to work and their earnings.

The emphasis on individual heterogeneity driving both earnings and selection has led to a large part of the segmentation debate focusing on whether studies have adequately controlled for this heterogeneity. For example Pratap and Quintin (2006) find that there is no evidence that earnings functions in the formal and informal sectors differ in equilibrium, using panel data from Argentina, once they adopt a semi-parametric propensity score matching approach that uses less restrictive assumptions than parametric approaches. Badaoui et al. (2008) find no evidence of an earnings premium for formal sector workers in South Africa once they use the panel dimension of their data to control for unobserved heterogeneity and account for only formal sector workers paying income tax¹. Botelho and Ponczek (2011) find a large premium for formal sector employees in Brazil using Ordinary Least Squares (OLS), but this becomes a small but statistically significant premium after they use a fixed effects regression to control for unobserved, time invariant heterogeneity.

It is clear that panel data have important advantages when looking at earnings differentials, since researchers are able to control for time invariant individual heterogeneity that may be the cause of differences in earnings across individuals. It is not often acknowledged, however, that the use of panel data

¹We discuss this paper in more detail in Section 2.3 below

also brings several challenges (Deaton 1997), and it is even more rare for these challenges to be addressed. The studies by Pratap and Quintin (2006) and Badaoui et al. (2008) mention neither measurement error nor attrition. Botelho and Ponczek (2011) emphasise that measurement error and attrition can influence the results obtained using panel data in their study of earnings differentials between the formal and informal sectors of the Brazilian labour market, but the only measurement error they deal with is that related to mismeasurement by respondents reporting incomes for other members of the household, for which a very simple test is presented. The authors do not deal with the more fundamental issue that measurement error in the independent variables in panel data can bias fixed effects regression coefficients towards zero (Deaton 1997), an issue that we take up below. Likewise, although attrition is addressed in Botelho and Ponczek (2011), it is dealt with only by noting that the observable characteristics of those who subsequently exit the sample are similar to those who stay in the sample. We discuss the results of attempting to correct for attrition and measurement error in section 4 below.

Thus far we have mentioned only the literature arguing that earnings differentials or their absence are evidence for or against segmentation. Testing for earnings differentials as an indication of whether the labour market is segmented is not uncontroversial, however. Maloney (1999) argues that heterogeneity in non-pecuniary aspects of employment is an important determinant of job satisfaction and also that the key implication of segmentation is a lack of movement between the informal and formal sectors, rather than earnings differentials. Despite this there is still active debate about the nature and causes of earnings differentials between different types of employment and we focus on this in this paper.

2.2 Beyond a formal/informal dichotomy

In much of the segmentation literature just reviewed the focus is fairly narrow, with researchers exploring whether there are earnings differentials between the formal and informal sectors. In some research the distinction between the formal and informal sectors is based on the ILO definition of informality, as defined in Hussmans (2004), ie whether the individual is registered or the firm the individual works for is registered (cf Badaoui et al. (2008)). In other research informality is defined as whether the individual receives benefits from their employer (Pratap and Quintin 2006) or has the correct registration card required in Brazil (Botelho and Ponczek 2011). However it is precisely defined, a binary dichotomy between the formal and informal sectors is helpful in exploring how the regulatory environment affects the earnings of those in the formal sector, but does little to enlighten our understanding of how other labour market institutions affect earnings in both the formal and informal sectors. What of the role of trade unions? How large is the public sector and are earnings in it high relative to the private sector? If there are multiple segments within the informal sector (Fields 2005), can we describe them? How does own account work compare to wage employment in small firms? These questions seem to have fallen

by the wayside in the debates over whether the labour market is segmented.

The approach we have outlined also links to a literature that argues that formality is not simply a dichotomy between the formal and informal sectors. Chen (2006) has argued that there is actually a continuum of the formality of jobs, with some being more formal than others. In this vein Günther and Launov (2012) argue that there is heterogeneity within the informal sector in Ivory Coast, and that the data they use suggests there are two distinct segments within informal employment, each of which pays less than the formal sector. Falco et al. (2011) also argue against a binary dichotomy, suggesting that enterprise size is a key determinant of formality in both Ghana and Tanzania.

In different ways the papers reviewed above suggest that the labour market is more complex than a description of a simple dichotomy between the formal and informal sectors. Our approach in this paper is similar in spirit. Partly this is driven by data considerations, since the data we use do not enable us to classify individuals as working for a registered enterprise along the lines of the most recent ILO definition of informality (Hussmans 2004). But we emphasise important distinctions within the formal sector, as well as comparing earnings between jobs that are more and less formal. Unionised, public sector work in South Africa could be considered highly formalised, unionised regular work in the private sector less so and self-employment or casual employment even less formal. The segmented labour market hypothesis suggests that jobs with higher levels of formality would be higher paying than less formal jobs, even after controlling for observed and unobserved ability. We test this hypothesis in our empirical analysis below, in addition to exploring the importance of both observed and unobserved human capital. In the following section we review the existing literature on the role labour market institutions play in determining earnings in South Africa.

2.3 The impact of institutions on the South African Labour market

There is much research focusing on the institutions influencing labour market outcomes in South Africa, particularly the impact of trade unions, the public sector and bargaining councils. Bargaining Councils are legally mandated groupings of firms and trade unions that set minimum wages for each industry, with the existing research suggesting that these raise wages and decrease employment (Magruder (2009), Borat et al. (2007)). Butcher and Rouse (2001) show that the earnings premium for trade union members was not as high in the 1990s as earlier estimates had suggested (Moll 1996), whilst Heintz and Posel (2008) show that public sector workers earn 40 % more than formal private sector workers after controlling for a range of observable characteristics. All of the work cited in this paragraph is cross sectional in nature, however.

A few studies specifically set out to investigate whether formal sector employment, usually defined as those working in or owning registered enterprises, is associated with a substantial earnings premium in South Africa. Heintz and

Posel (2008) find large premia for formal sector workers using cross-sectional data from the 2004 Labour Force Survey.

Results from cross-sectional data that suggest the existence of formal sector earnings premia may actually be explained by individual differences in unobserved ability or unobserved school quality, however. The only study that takes into account unobserved individual heterogeneity using South African data is Badaoui et al. (2008). The authors use the Labour Force Survey panel to show that there is no formal sector premium for black, male employees once unobservable differences between individuals are taken into account. We noted above that one problem with this study is that it does not account for the additional problems panel data bring. In addition, Badaoui et al. (2008) define a very narrow field of investigation. The authors look only at private employees, excluding public sector workers and the self-employed from their study. The authors thus only attempt to explore whether there is an earnings premium for employees in registered private enterprises, compared to private unregistered enterprises. In finding no formal sector premium this paper tells one little about *how* this result is obtained or the institutions or regulations that have contributed to this situation. In contrast, our work in this paper attempts to explore the effects of different institutions on wage setting both *within* formal employment and in less formal types of employment, using the KIDS data, which we now describe.

3 Description of the KIDS Data and Survey Methodology

This section provides a brief description of the KwaZulu-Natal Income Dynamics Study (KIDS), a 3 wave panel conducted in 1993, 1998 and 2004 in KwaZulu-Natal province, which was the largest province by population when the survey was undertaken. As noted above we confine our analysis to black individuals only.

3.1 Describing Employment, Earnings and Education

The KIDS data allow us to explore the effects of public sector wage setting and unionisation on earnings, which can be argued to be two important dimensions of formality. We thus describe the differences in earnings between six mutually exclusive employment types that can be identified in the KIDS: four types of regular employment (private non-union, private union, public union and public non-union) as well as casual employment and self-employment.

The structure of employment in South Africa reflects a major difference between it and most other African countries: self-employment in informal enterprises and other types of informal employment make up only a small part of total employment (Kingdon and Knight 2004). This difference is evident in Table 1 for the first two waves of the KIDS data, where regular employment predominates. By 2004, however, casual employment had become the most common form of employment, with a dramatic decline in regular employment,

possibly reflecting non-random attrition from the panel. Table 2 shows that employment decreased by about ten percent between 1993 and 2004, and that there was a corresponding increase in the measured broad unemployment rate².

Table 1 also shows the large hourly earnings differences between regular, casual and self-employment that we highlighted at the start of this paper³. The median wage for public sector workers in unions was 17 times higher than earnings in self-employment in 2004, a gap which had increased three-fold since 1993. This increase in the earnings gap over the 11 year period covered by the surveys was generated by a decline in earnings by a third for the self-employed and an increase of roughly 65% for unionised public sector employees. There were substantial increases in real wages between 1993 and 1998, as noted in Cichello et al. (2005), with the largest percentage increases for public sector workers. This may reflect the raising of the wages of black individuals in the public sector to the level of their white counterparts as a result of the new post-1994 government, or the effects of the formation of the Public Service Coordinating Bargaining Council, which occurred in 1997 (Bhorat et al. 2007). Table 1 also shows that between 1998 and 2004 real wages remained constant or decreased slightly.

Table 4 shows occupations by employment type⁴. It is clear that some types of employment are more likely to be associated with certain types of occupations. Public sector employees are more likely to be in professional occupations, whilst regular private employees are more likely to be labourers or in production jobs. This is some indication that part of the explanation for the vast differences in earnings across employment types must be found in human capital and occupational choice. We explore this in more detail below.

Significant educational expansion between 1993 and 2004 is evident in the KIDS data, shown in Table 3. Public sector employees have the highest levels of education across the 3 waves, whilst individuals in self-employment have the lowest. Large earnings differences across the sectors seem to be correlated with large differences in educational attainment, consistent with human capital theory. This will be given further attention in the subsequent analysis. We turn next to a discussion of attrition in the KIDS panel.

3.2 Attrition

In any panel survey attrition is a potential cause for concern, as it can generate attrition bias in parameter value estimates. The KIDS methodology involved attempting to track those households that moved, although not all households were found. At least one core member of the 1132 households first interviewed in 1993 was successfully re-interviewed in 721 households in 2004, representing household attrition of 36 percent over eleven years. It was also possible for

²The number of employed and the number of earners in the sample differs because some individuals did not provide earnings measures. We have not imputed earnings where these were missing.

³The Appendix gives details about how the earnings measure was constructed.

⁴The self-employed were not asked questions about their occupation.

individuals to exit from households that were re-interviewed. Of the 8258 individuals in households that were interviewed in 1993, 5397 were re-interviewed in 2004. Whilst these are large rates of attrition May et al. (2007, pg. 638) show that they are comparable to attrition rates in panel studies in other developing countries such as the LSMS Cote d'Ivoire panel survey conducted in the late 1980s and the Lima LSMS from Peru and argue that KIDS “continues in most of its important demographic characteristics to remain broadly representative of the population of the province.”

Table 5 show some correlates of household attrition. In both 1998 and 2004 it seems that smaller households are more likely to exit the sample. Households with higher per capita income seem more likely to exit in 2004, but not in 1998. Table 6 shows some correlates of individual attrition (all individuals, including those who exit the sample due to household attrition, older than 15). Individuals that exit are older, more likely to be male, come from larger households and have slightly higher levels of education. Whether the individual is a “core” member or not is an important predictor of attrition in both 1998 and 2004: this is because the survey methodology meant core household members were tracked if they moved, whilst other household members were not. Table 6 shows similar characteristics for the sub-sample of earners, which we use in the next section. In this sample men and those who are married seem more likely to exit. Younger earners also seem more likely to exit the sample.

Having given a description of attrition in the KIDS we now turn to explaining wage differentials.

4 Explaining Wage Differentials in South Africa

4.1 Can Human Capital Theory Account for Earnings Differentials?

In this section we explore the extent to which the competitive model, within a human capital framework, can explain the large average wage differentials across employment of different levels of formality documented above. To begin we estimate earnings functions using Ordinary Least Squares (OLS):

$$y_{it} = \gamma + x'_{it}\beta + \varepsilon_{it}. \quad (1)$$

x_{it} and β are $K \times 1$ vectors, and we assume the ε_{it} are iid $[0, \sigma^2]$. Table 8 shows basic OLS earnings functions for the 3 waves of the panel, with the log of real hourly wage being the dependent variable⁵. Missing earnings or hours data, as a result of individuals refusing to answer or because the respondent did not know, are not imputed. Around 4 percent of the sample had missing earnings and less than 1 percent had missing hours worked. The education variables

⁵The interpretation of the regression coefficients is the percentage change in the hourly wage for a unit change in a regressor. For dummy variables the percentage effect of a change from zero to one is calculated by $e^\beta - 1$. Table 7 in the Appendix to this paper shows basic summary statistics for the variables included in the regressions.

used are dummies indicating whether the individual had some primary, completed primary, some secondary, matric (completed secondary) or some tertiary education. The omitted category is no education.

Across all three waves the results show that there are large returns to education and labour market experience. The strongly convex returns to education are consistent with other findings for South Africa (Keswell and Poswell 2004) and for other African countries (Bennell 1996), and contradict the assertion of Psacharopoulos (1994) that returns to education are generally concave in developing countries.

The vast raw wage differences across various dimensions of formality, noted in section 3, are substantially reduced in the OLS earnings functions for each of the 3 waves, conditioning on a range of covariates. The six-fold raw difference in average earnings over the 3 waves of KIDS between the highest and lowest paying sectors is reduced to roughly a four-fold difference when controlling for observable human capital. Interestingly, most of this occurs through reduced differentials within the private sector, with the large raw public sector premium hardly declining. The union premium in the private sector is estimated to be around 60%, but is roughly 30% when we control for industry and occupation, shown in the last column of Table 8. This is broadly in line with the analysis of Butcher and Rouse (2001)⁶. In section 3 we showed that public sector workers are more likely to be in professional occupations. The results from the last column of Table 8 suggest that this is not the cause of the public sector premium, which remains when controls for industry and occupation are included.

4.2 Modeling Unobserved Heterogeneity

OLS estimation fails to allow for factors such as unobserved ability and preferences for different job characteristics. Panel estimators can be used to control for unobserved individual time invariant heterogeneity. It is then possible to estimate the change in the wage as an individual moves between union and non-union employment, or into or out of the public sector. There is a significant amount of movement between sectors, as shown in Tables 9 and 10, which is required for identification of differences in earnings across different types of employment using panel estimators. In particular, anticipating our results below, there is extensive movement out of, and into, private unionised employment.

We use fixed effects and first and second difference estimators in our estimation. The fixed effects model is

$$y_{it} = \gamma + \alpha_i + x'_{it}\beta + \varepsilon_{it}. \quad (2)$$

The α_i terms are the individual specific effects. The fixed effects estimator is obtained by subtracting the time averaged model from the original model, giving

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i). \quad (3)$$

⁶With the KIDS data we are only able to explore the effect of union membership and not on the broader role unions play in raising minimum wages for all workers through the bargaining council system that we discussed in Section 2.

Wooldridge (2002) shows that this is a consistent and unbiased estimator if either $N \rightarrow \infty$ or $T \rightarrow \infty$ and assuming strict exogeneity of the ε_{it} , ie that $E[\varepsilon_{it}|x_{it}, \alpha_i] = 0$. Alternatively one can estimate the fixed effects model using the first difference estimator, subtracting the model lagged one period, to obtain

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})'\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}). \quad (4)$$

This is also unbiased and consistent under the assumptions above, but is less efficient than the fixed effects estimator when ε_{it} are serially uncorrelated (Wooldridge 2002). Longer differencing periods can reduce the effects of attenuation bias due to measurement error and help determine how serious the bias is in the first differenced results (Griliches and Hausman 1986). We explore this further below.

4.3 The effects of unobserved heterogeneity

Table 11 reports results from the fixed effects estimator, as well as the first and second difference estimators. Second differencing is the maximum difference allowed by the data, since KIDS is a three wave panel. The first column shows the results of the fixed effects estimator. They suggest that a substantial premium remains for those in public sector employment, relative to private regular employment. Those in unionised, public sector employment earn 91 percent more than those in private, non-union employment, controlling for time invariant unobservables and time varying observables. Private sector, regular employment is more lucrative than casual or self-employment, with those in casual employment earning 27 percent less, and those in self-employment 9 percent less than those in regular, private, non-union employment, although the self-employment coefficient is not significantly different from zero at the 5 percent level.

There is also a statistically significant⁷, two and a half fold difference between the highest and lowest paying sectors. Controlling for unobserved heterogeneity further reduces the differentials within the private sector, such that the differences are not significant in the fixed effects regression. The public sector premium, however, is hardly reduced at all. The results from the first difference estimator are similar to those from the fixed effects estimator and also indicate that a large premium remains for those in public sector employment, relative to private, regular employment, but that the private sector union premium is much lower than in the OLS regressions and insignificant.

The small union premium in the private sector is an important finding, as it suggests that the larger union premium in the cross section results may actually be a quality effect, and that union members earn higher wages than non-union members in private employment because they are (unobservably) more productive. It is also likely that firms are responding to unionisation, and having to pay higher wages, by increasing their capital stock and lowering their labour usage but hiring higher quality workers necessary to utilise the

⁷Table 12 contains the results of pairwise F tests for the equality of the coefficients on the different types of employment for the fixed effects regression.

increased capital. This finding suggests that union activity is changing the skill composition of the workforce.

We noted above that there is substantial movement into and out of union membership, which should allow for the identification of the union effect in both the fixed effects and differenced estimators. This requirement for movement means that panel estimators rely on those observed more than once to identify the regression coefficients. Those only observed once are excluded. As a simple robustness check, it is helpful to determine whether the pooled cross section results also hold for the sub-sample of individuals observed more than once. The last column in Table 15 shows the key OLS results are indeed similar in the reduced sample, with earnings increasing as the degree of formality of employment increases. We now explore some other potential sources of bias in our results.

4.4 Measurement Error using Panel Data

Measurement error is of concern in any data, and this can be exacerbated in panel data (Deaton 1997). A well known result from econometric theory is that measurement error in an independent variable results in attenuation bias, meaning that coefficients of the independent variables are biased towards zero. Following Cameron and Trivedi (2005, pg. 905), the effects of measurement error in panel data can be illustrated using a model with a scalar regressor:

$$y_{it} = \alpha_i + x_{it}^* \beta + \varepsilon_{it}. \quad (5)$$

α_i is the individual unobserved fixed effect, x^* is measured with error and we observe x , with $x_{it} = x_{it}^* + v_{it}$. If we use a first difference estimator then

$$\Delta y_{it} = \beta \Delta x_{it}^* + \Delta \mu_{it} \quad (6)$$

$$= \beta \Delta x_{it} + \Delta \mu_{it} - \beta \Delta v_{it}. \quad (7)$$

If we define $\rho = \text{Cor}(x_{it}^*, x_{i,t-1}^*)$, then it can be shown that

$$plim \hat{\beta} = \beta + (plim \frac{1}{N} \sum_{i=1}^N \Delta x_{it}^2)^{-1} plim \frac{1}{N} \sum_{i=1}^N (\Delta x_{it} \Delta \mu_{it} - \beta \Delta x_{it} v_{it}) \quad (8)$$

$$= \beta - \frac{\beta \sigma_v^2}{(1 - \rho) \sigma_{x^*}^2 + \sigma_{v_v}^2}. \quad (9)$$

This inconsistency is larger than in the cross sectional case when $\rho > 0$. It is also clear from equation (9) that as $\rho \rightarrow 1$ the inconsistency becomes large. KIDS has longer periods between waves than many other panel surveys, with 5 years between the first 2 waves and 6 years between the second and third waves. This would help to alleviate attenuation bias if the dependent variable changes by a larger amount, compared to panels with shorter periods between waves, as this would lower the value of ρ .

We noted above the large decline in the union premium when using the fixed effects and first difference estimators and gave a possible economic explanation. This could also be attributed to measurement error in sector of employment, however, and the resultant attenuation bias. As a further robustness check, longer differencing periods can reduce the effects of attenuation bias and help determine how serious the bias is in the first difference results (Griliches and Hausman 1986). The KIDS is a three wave panel, and hence the maximum differencing possible allows for a second difference estimator. If measurement error was driving our results then we would expect to see larger coefficients in the second difference estimator than in the first difference. Table 11 shows the sector coefficients in the second difference results are generally not significant, probably as a result of the small sample, but that they are not universally larger than the fixed effects and first difference coefficients. In fact, the private sector union premium is *lower* in the second difference regression than the first difference regression, and is in fact negative, implying that our low estimate of the premium is unlikely to be the result of measurement error⁸.

4.5 Interrogating the Public Sector Earnings Premium

In order to obtain consistent estimates of the differences in earnings across sectors when accounting for time invariant heterogeneity we have thus far assumed that movement between sectors is exogenous. Similar assumptions are common in much of the literature, including the seminal work of Abowd et al. (1999), who use matched firm and worker data to decompose earnings into components due to observed and unobserved characteristics of firms and workers, but assume that movement of workers between firms is exogenous.

In this section we use insights from the treatment effects literature to show that movement is endogenous if the return to different sectors differs by individual and provide some initial evidence that this is the case in the KIDS data and that this affects the estimation of the public sector premium. We do not attempt to correct for endogenous movement as we do not have instruments that explain movement but not earnings. Falco et al. (2011) use system-GMM estimators to estimate the effect of endogenous movement on sectoral differentials in Tanzania and Ghana, although they do caution against reading too much into the system-GMM results given that they use a three wave panel. Unfortunately the small sample of individuals observed with earnings in all three periods does not allow us to replicate this analysis using the KIDS data.

In the treatment effects literature heterogeneity in treatment is an important part of the endogeneity problem researchers must overcome when estimating treatment effects (Heckman 2001). In this world earnings may actually be determined by the following model:

$$y_{it} = \gamma + \alpha_i + x'_{it}\beta_i + \varepsilon_{it}, \quad (10)$$

⁸Butcher and Rouse (2001) show that ignoring controls for industry *inflates* the union premium. Our estimates do not control for industry and can thus be thought of as the *maximum* estimate of the union premium, assuming we have controlled for other potential biases.

where x is a vector of the different sectors and the returns to these sectors are now assumed to be individual specific. This means that in first differences we estimate an average of the returns to each sector across all individuals who change sectors:

$$\Delta y_{it} = \bar{\beta} \Delta x_{it} + \Delta \varepsilon_{it} + (\beta_i - \bar{\beta}) \Delta x_{it}. \quad (11)$$

In the language of the treatment effects literature we are attempting to estimate the average treatment effect on earnings of being “treated” by being in a particular sector. The error term in equation (11) does not satisfy the strict exogeneity assumption that is required for consistency, however, and implies that movement between sectors will not be exogenous, and hence that we may not consistently estimate $\bar{\beta}$ if there is individual heterogeneity in β .

We provide some indication that there may be individual heterogeneity in β , and thus endogeneity of movement between sectors, by focusing on the large public sector earnings premium we found even after controlling for unobserved individual time invariant heterogeneity. The second column in Table 13 uses a first difference estimator but excludes those who move out of the public sector. This means the public sector premium is being estimated using only those who move into the public sector. These results are similar to the full sample, shown in column 1. Column 3 shows the results excluding those who enter the public sector, which are radically different, with no significant public sector premium⁹. Columns 4-6 repeat the analysis combining unionised and non-unionised public sector workers, and the same result obtains.

Despite having controlled for individual unobserved heterogeneity, α_i , using the first difference estimator, our results suggest there are still differences in relative returns to public sector employment. What could explain this result? If individual productivity is captured by α_i and controlled for in the panel estimators then the coefficient on the public sector dummy can be thought of as the relative returns to that sector. In this case it is not inconceivable that low productivity individuals have higher returns in the public sector, relative to other sectors, if public sector pay is not as strongly correlated with ability. This scenario would also require that individuals of all productivity levels have some chance of gaining entry to public sector employment (perhaps because of poor screening of applicants or nepotism) and that being fired from the public sector is very difficult. These are not unrealistic assumptions in the South African context.

Under these assumptions low ability workers may manage to enter public employment, knowing that finding better paying employment in the private sector is not possible. As a result these workers would stay in the public sector for as long as possible, earning a premium compared to their previous earnings in private sector employment, which is what we estimate when excluding those who move out of the public sector. High ability workers may have lower relative returns to the public sector if public sector earnings are less correlated with ability. This would mean their earnings could be matched or exceeded if they received a job offer in the private sector, implying they would be the only workers

⁹A similar analysis for the union effect in the private sector produces no asymmetric effect.

who leave public sector employment and explaining why when we exclude those who move into the public sector we do not find a public sector premium.

4.6 Further Robustness Checks

We have noted that non-random attrition may be affecting our results. In the Appendix we set out attrition correction methods and results for the OLS and differenced regressions. These involve the use of inverse probability weighting in the case of OLS and a Heckman selection correction for the differenced results (Wooldridge 2002). Both require fairly strong parametric assumptions for identification. They do not make dramatic differences either to the OLS or differenced results, and provide some indication that attrition is not biasing our results.

A simpler final check on the effects of attrition on the results that correct for unobserved heterogeneity is to estimate the first difference model using the first two waves only and then the last two waves only, shown in the final two columns of Table 16. The results are similar to the uncorrected and attrition corrected results.

As a final set of robustness checks we report OLS and fixed effects regressions dividing the sample by gender. Table 14 shows that the key results when separating the sample into men and women are preserved. The ranking of sectors is roughly the same for the full sample, although women are relatively worse off in self-employment. The union premium in the private sector is again dramatically lowered when estimated using the fixed effects panel estimator for both men and women.

5 Conclusion

The KIDS panel enables us to explore whether the large earnings differentials between jobs that differ across various dimensions of formality are the result of unobservable differences between individuals or due to the effects of labour market institutions. Once we use panel estimators, to control for unobservable time invariant heterogeneity, the differential between the highest and lowest paid sectors is reduced to a two and half fold difference. Attrition and measurement error can bias results when using panel data but we have shown our results are robust to these concerns.

The only substantial earnings premium found is for those in public sector employment, the most formal of the employment categories we consider. Interestingly, we find that this premium is driven by those entering the public sector, rather than those leaving it, and we have outlined a very simple model of ability driving movement into and out of the public sector to explain this result.

We also find that the union premium in the private sector is reduced substantially when controlling for individual heterogeneity, suggesting that firms in which unions are active attempt to mitigate higher wages by selecting employees more carefully. Previous research on the role of trade unions in South Africa

concentrated on whether unions generate a premium for their members. Our research has shown the situation is more complex, since unionised workers are of unobservably higher quality. This suggests that an important effect of union activity is to change the skill composition of the work force.

Our conclusions in this paper about the role of trade unions and the public sector suggests that concentrating on a binary dichotomy between the formal and informal sectors is often unhelpful in exploring how other institutions affect the labour market, and that this approach often precludes analysis of important labour market institutions. Our results suggest that the explanation for the large earnings differences across different types of employment in South Africa is not found only in human capital or only in institutions that disrupt the equilibrium in the labour market. Instead, we have shown that human capital explains much of the earnings differentials within the private sector but cannot explain the large public sector premium.

6 Appendix

6.1 Construction of the earnings and hours worked variable

The earnings variable is constructed from net earnings for those who reported paying tax and gross earnings for those who reported not paying tax. For those in self-employment the questions were about the total hours and profits in the business, and households were allowed to list up to three members involved in the business, but there was no question asking how much time each member worked in the business. We have assumed an equal share of time and profit for each individual listed as helping with the business. The hours worked variable is constructed using a question on the number of hours worked in an average day for those in casual and regular employment.

6.2 Tracking of Households and household types

The 1998 KIDS survey attempted to track households, including those that moved between waves and those that split up, interviewing all members of any households that contained core members of the 1993 households. Core members were those that satisfied any of the following four criteria: (i) A self-declared head of household (from 1993)

(ii) Resident spouse/partner of self-declared head of household (from 1993)

(iii) Resident member in a three-generation household (from 1993) and all of the following were true:

child, child-in-law, or niece/nephew of self-declared head;

at least 30 years old;

have at least one child resident in the household

(iv) Resident spouse/partner of individual satisfying criterion (iii)

A similar procedure was followed in the third wave in 2004 although the panel was refreshed by also tracking “next generation” households, defined as those containing adult children of core members, who themselves have children.

6.3 Addressing Attrition

We model attrition at an individual level and then use two different attrition correction methods to correct our results. The pooled cross section regressions are corrected using Inverse Probability Weighting (IPW) and the difference estimator regressions are corrected using a Heckman selection correction, following Wooldridge (2002).

6.3.1 Modeling Attrition

We model attrition as a binary variable s_{it} that takes a value of one if the individual exits the sample at time t after appearing in the sample at time $t - 1$

and is zero otherwise. s_{it} is a function of individual, household and community characteristics that are represented by the vector w_{it} :

$$s_{it} = 1[w_{it}\delta + v_{it} > 0]. \quad (12)$$

Tables 17 and 18¹⁰ shows the results of 3 probit models of individual attrition. The first column uses 1993 characteristics of individuals themselves, as well as characteristics of their households and communities, to predict whether individuals exit the sample in 1998, the second wave of the panel. The second uses the same characteristics to predict attrition in 2004. The third column uses characteristics of all those present in 1998 (including new household members not present in 1993) to predict attrition in 2004.

The results suggest that both elderly household members and the very young, the omitted category, are least likely to exit. Attrition is also found to be less likely for resident household members, for males and for those in larger households. Those in private sector regular unionised employment are also less likely to exit the sample. No individual in any other category of employment or the unemployed were significantly more likely to exit compared to the omitted category, which was those not in the labour force.

6.3.2 Correcting for Possible Attrition Bias

In Section 4 individual earnings were modeled as

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}. \quad (13)$$

In correcting for possible attrition bias in the pooled cross section Inverse Probability Weighting relies on the so-called ‘‘ignorability’’ assumption (Wooldridge 2002):

$$P(s_{it} = 1|y_{it}, x_{it}, w_{i1}) = P(s_{it} = 1|w_{i1}). \quad (14)$$

Inverse Probability Weighting requires using the predicted probabilities of exit from the sample from a model of attrition, for example the attrition probit estimated above, and then weighting observations so that those that have a higher predicted probability of attriting, but nevertheless stay in the sample, are given a higher weight. These results are shown in the first column of Table 15, and are not dramatically different from the uncorrected results. An optimistic interpretation of the results is that attrition is not a serious concern. It could also indicate, however, that the ignorability assumption has been violated, given the fairly low predictive power of the variables in the attrition model.

In correcting for possible attrition bias in the differenced estimation, we follow Wooldridge (2002) and assume joint normality of $\Delta\varepsilon_{it}$ and v_{it} , which are the errors in the attrition and differenced earnings functions respectively, strict exogeneity of x_{it} and that selection does not depend on Δx_{it} once w_{it} has been controlled for. It can then be shown that

$$E(\Delta y_{it}|\Delta x_{it}, w_{it}, s_{it} = 1) = \beta\Delta x_{it} + \rho\lambda(w_{it}\delta), \quad (15)$$

¹⁰The second table reports the last half of the regression results from the first table due to the large number of explanatory variables.

where $\lambda(w_{it}\delta)$ is the inverse mills ratio from the attrition probit.

Table 16 shows the first and second differenced equations corrected for attrition. The results again do not differ dramatically to the uncorrected results. The key element of formality is again seen to be access to public sector employment, with union membership in the private sector not associated with a significant premium compared to non-union workers.

7 Acknowledgements

We thank Justin Sandefur and Rulof Burger as well as seminar participants at the 2009 CSAE conference in Oxford and the 4th IZA/World Bank Conference on Employment and Development in Bonn for helpful comments.

Table 1: MEDIAN HOURLY EARNINGS

	1993		1998		2004	
	<u>Median N</u>		<u>Median N</u>		<u>Median N</u>	
Employment Categories						
Public Union	13.90	61	24.35	116	23.39	138
	(9.73)		(56.12)		(77.80)	
Public non-Union	10.77	100	18.27	96	17.96	41
	(14.14)		(42.16)		(49.89)	
Private Union	9.94	168	12.15	152	11.44	66
	(10.90)		(32.89)		(38.19)	
Private non-Union	4.04	459	6.55	300	8.59	126
	(15.62)		(16.65)		(28.53)	
Casual Employment	2.19	91	5.16	167	3.70	299
	(4.58)		(18.70)		(8.73)	
Self Employment	2.32	156	3.44	88	1.02	66
	(29.11)		(33.57)		(5.26)	
Frequency	1081		961		789	

Standard Deviations in Parentheses. Earnings expressed in constant 2004 Rand. Average Rand/ US Dollar exchange rate was 3.26 in 1993, 5.52 in 1998 and 6.46 in 2004 (International Monetary Fund 2009)

Table 2: LABOUR FORCE PARTICIPATION AND UNEMPLOYMENT

	1993	1998	2004
Employed	1171	1065	1028
Labour Force Participants	1990	2118	1968
Unemployment Rate	41%	50%	48%

Source: Own calculations from KIDS.

Table 3: MEAN WEEKLY HOURS WORKED AND EDUCATIONAL ATTAINMENT (YEARS)

	1993		1998		2004	
	Hours	Educ	Hours	Educ	Hours	Educ
Employment Categories						
Public Union	43.90 (10.97)	10.36 (3.98)	37.36 (16.97)	11.2 (2.94)	38.62 (16.16)	12.4 (3.32)
Public non-Union	44.96 (14.83)	9.79 (4.13)	38.23 (18.44)	9.3 (4.30)	36.76 (18.49)	9.5 (4.51)
Private Union	47.24 (11.10)	7.51 (3.36)	43.15 (13.28)	8.38 (3.27)	45.36 (15.19)	8.34 (3.96)
Private non-Union	45.43 (15.57)	5.88 (3.85)	42.56 (14.89)	6.69 (3.95)	42.91 (18.76)	8.5 (3.76)
Casual Employment	38.18 (17.35)	5.25 (4.04)	34.74 (19.35)	7.57 (3.82)	40.78 (20.20)	7.97 (4.04)
Self Employment	35.46 (25.28)	5.62 (3.79)	48.16 (27.39)	6.16 (3.87)	39.05 (24.01)	8.28 (3.79)
Frequency	1081		961		789	

Standard Deviations in Parentheses. Source: own calculations from KIDS.

Table 4: OCCUPATION OF EMPLOYMENT BY EMPLOYMENT TYPE

Employment Types	Prof	Clerical	Farm	Labourer	Production	Services
	Public Union	186	26	1	27	52
Public non-Union	90	28	1	59	38	34
Private Union	20	19	47	147	178	36
Private non-Union	36	46	56	415	204	143
Casual Employment	39	28	0	216	127	158
Total	380	147	111	879	607	429

Source: Own calculations from KIDS.

Table 5: HOUSEHOLD ATTRITION DESCRIPTIVE STATISTICS

Correlates	1993		1998		2004
	Resurveyed	Attrited	Resurveyed	Attrited	Full Sample
HH Size	7.56 (4.17)	5.55 (3.46)	8.65 (4.54)	5.15 (3.22)	6.96 (4.07)
Per cap Income	178.25 (241.87)	186.27 (254.85)	291.88 (516.24)	523.64 (719.70)	459.14 (1280.90)
Rural	0.73 (0.44)	0.81 (0.39)	0.77 (0.42)	0.65 (0.48)	0.77 (0.42)
Own house	0.84 (0.36)	0.77 (0.42)	0.92 (0.27)	0.69 (0.47)	0.90 (0.30)
Frequency	964	175	752	248	1211

Standard Deviations in Parentheses. Source: own calculations from KIDS.

Table 6: INDIVIDUAL ATTRITION DESCRIPTIVE STATISTICS

	1993		1998		2004
	<u>Resurveyed</u>	<u>Attrited</u>	<u>Resurveyed</u>	<u>Attrited</u>	<u>Full Sample</u>
Full Sample					
Age	23.31 (18.94)	22.50 (16.14)	26.53 (19.30)	24.34 (16.45)	24.26 (18.68)
Male	0.47 (0.50)	0.47 (0.50)	0.46 (0.50)	0.49 (0.50)	0.47 (0.50)
Married	0.24 (0.43)	0.25 (0.43)	0.19 (0.39)	0.17 (0.38)	0.15 (0.36)
Year of Educ	4.62 (4.06)	5.14 (4.17)	5.53 (4.25)	6.17 (4.31)	6.38 (4.42)
Frequency	6334	1924	6507	2688	8318
Earner Sample					
Age	37.80 (11.97)	34.15 (11.69)	39.89 (11.40)	37.57 (10.69)	38.13 (11.16)
Male	0.52 (0.50)	0.48 (0.50)	0.47 (0.50)	0.55 (0.50)	0.49 (0.50)
Married	0.60 (0.49)	0.57 (0.50)	0.49 (0.50)	0.53 (0.50)	0.42 (0.49)
Year of Educ	6.54 (4.18)	6.62 (4.03)	7.54 (4.09)	8.10 (3.93)	8.93 (4.25)
Frequency	737	239	526	354	750

Standard Deviations in Parentheses. Source: own calculations from KIDS.

Table 7: EARNER SAMPLE SUMMARY STATISTICS

	1993	1998	2004
Age	36.90 (12.00)	38.96 (11.17)	38.13 (11.16)
Male	0.51 (0.50)	0.50 (0.50)	0.49 (0.50)
Married	0.60 (0.49)	0.50 (0.50)	0.42 (0.49)
No Educ	0.16 (0.37)	0.09 (0.29)	0.07 (0.25)
Some Primary	0.30 (0.46)	0.25 (0.43)	0.19 (0.39)
Completed Primary	0.10 (0.30)	0.09 (0.29)	0.07 (0.26)
Some Secondary	0.31 (0.46)	0.33 (0.47)	0.34 (0.47)
Completed Secondary	0.09 (0.28)	0.19 (0.39)	0.20 (0.40)
Frequency	976	880	750

Standard Deviations in Parentheses. Source: own calculations from KIDS.

Table 8: OLS EARNINGS FUNCTIONS

	93	98	04	pooled	IO
	(1)	(2)	(3)	(4)	(5)
Age	0.064 (0.018)***	0.039 (0.023)*	0.065 (0.019)***	0.062 (0.01)***	0.056 (0.01)***
Age ²	-0.0006 (0.0002)***	-0.0003 (0.0003)	-0.0007 (0.0002)***	-0.0006 (0.0001)***	-0.0006 (0.0001)***
Incomplete prim educ	0.26 (0.091)***	0.527 (0.166)***	0.222 (0.193)	0.302 (0.085)***	0.227 (0.073)***
Complete prim educ	0.442 (0.117)***	0.791 (0.178)***	0.382 (0.258)	0.504 (0.107)***	0.4 (0.095)***
Incomplete sec educ	0.691 (0.104)***	1.073 (0.18)***	0.558 (0.211)***	0.738 (0.104)***	0.547 (0.076)***
Complete sec educ	0.884 (0.18)***	1.422 (0.216)***	0.822 (0.219)***	1.040 (0.134)***	0.73 (0.108)***
Tertiary educ	1.515 (0.154)***	1.731 (0.255)***	1.429 (0.253)***	1.537 (0.127)***	1.063 (0.116)***
Married	0.178 (0.075)**	-0.131 (0.085)	0.113 (0.09)	0.061 (0.047)	0.053 (0.047)
Male	0.439 (0.075)***	0.316 (0.069)***	0.124 (0.076)	0.299 (0.042)***	0.252 (0.039)***
Public Union	0.615 (0.16)***	0.977 (0.116)***	0.947 (0.142)***	0.896 (0.09)***	0.69 (0.081)***
Public non-Union	0.444 (0.111)***	0.808 (0.144)***	0.869 (0.177)***	0.664 (0.083)***	0.486 (0.069)***
Private Union	0.517 (0.109)***	0.504 (0.134)***	0.456 (0.141)***	0.506 (0.088)***	0.278 (0.067)***
Casual Employment	-0.452 (0.13)***	-0.147 (0.13)	-0.325 (0.093)***	-0.283 (0.065)***	-0.246 (0.062)***
Self employment	-0.449 (0.159)***	-0.144 (0.193)	-0.849 (0.177)***	-0.463 (0.099)***	-0.375 (0.098)***
Const.	-0.605 (0.338)*	-0.080 (0.521)	-0.141 (0.449)	-0.537 (0.227)**	-0.353 (0.222)
Obs.	976	880	750	2606	2606
R ²	0.347	0.36	0.428	0.381	0.421

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Standard errors are in parentheses and adjusted for intra-cluster correlation. The dependent variable is the log of hourly net income. All variables are dummies except age variables. Reference categories are no education and private non-union employment. The IO column adds industry and occupation controls to the pooled data.

Table 9: EARNER MOVEMENT 93-98

	regprivatenonunion	regpubunion	regpubnonunion	regprivateunion	casearner	selfearner	Total
regprivatenonunion	101	10	6	38	29	12	196
regpubunion	3	12	8	2	2	1	28
regpubnonunion	8	19	12	6	7	3	55
regprivateunion	24	15	7	31	14	7	98
casearner	7	1	6	3	11	4	32
selfearner	8	4	1	5	10	38	66
Total	151	61	40	85	73	65	475

57

An individual's 1993 sector is on the left and 1998 sector is on the top. Source: own calculations from KIDS.

Table 10: EARNER MOVEMENT 98-04

	regprivatenonunion	regpubunion	regpubnonunion	regprivateunion	casearner	selfearner	Total
regprivatenonunion	20	6	1	10	40	6	83
regpubunion	3	30	8	5	2	1	49
regpubnonunion	3	15	1	3	6	0	28
regprivateunion	8	4	2	11	10	6	41
casearner	7	1	3	4	18	2	35
selfearner	1	2	1	0	13	30	47
Total	42	58	16	33	89	45	283

An individual's 1998 sector is on the left and 2004 sector is on the top. Source: Own calculations from KIDS.

Table 11: FIXED EFFECTS AND DIFFERENCED EARNINGS FUNCTIONS

	Fixed Effects	First Difference	Second Difference
	(1)	(2)	(3)
Age	0.074 (0.035)**	0.049 (0.029)*	-.451 (0.316)
Age ²	-.0009 (0.0004)**	-.0007 (0.0004)*	0.004 (0.003)
Public Union	0.622 (0.161)***	0.461 (0.149)***	0.385 (0.318)
Public non-Union	0.47 (0.153)***	0.344 (0.164)**	0.745 (0.386)*
Private Union	0.167 (0.091)*	0.125 (0.12)	-.030 (0.211)
Casual Employment	-.199 (0.108)*	-.255 (0.135)*	-.373 (0.292)
Self employment	-.128 (0.172)	-.147 (0.186)	-.266 (0.257)
1998 Year Dummy	0.393 (0.105)***	0.534 (0.163)***	
2004 Year Dummy	0.263 (0.171)		
Const.	0.32 (0.739)	-.107 (0.162)	-.826 (0.263)***
Obs.	1154	628	126
R ²	0.105	0.084	0.101

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Standard errors are in parentheses and adjusted for intra-cluster correlation. The dependent variable is the log of hourly net income. Reference employment category is private non-union employment.

Table 12: F TESTS OF EQUALITY OF SECTORAL COEFFICIENTS FROM FIXED EFFECTS REGRESSION

Labour Market Status	regpubNU	regprivU	casual	selfemp
regpubU	0.97 (0.33)	8.95 (0.004)	28.27 (0.000)	14.57 (0.000)
regpubNU		4.53 (0.037)	13.91 (0.000)	8.72 (0.005)
regprivU			9.18 (0.004)	3.15 (0.081)
casual				0.17 (0.679)

Notes: p values are in parenthesis. NU indicates non-union and U indicates union.

Table 13: FIRST DIFFERENCED EARNINGS FUNCTIONS EXCLUDING MOVERS
 INTO AND OUT OF PUBLIC SECTOR EMPLOYMENT

	Orig	Exclout	Exclinto	Pub	PubExclout	PubExclinto
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.049 (0.029)*	0.045 (0.028)	0.044 (0.032)	0.048 (0.029)*	0.046 (0.028)	0.043 (0.032)
Age ²	-.0007 (0.0004)*	-.0007 (0.0004)**	-.0005 (0.0004)	-.0007 (0.0004)*	-.0007 (0.0004)**	-.0005 (0.0004)
Public Union	0.461 (0.149)***	0.423 (0.176)**	0.155 (0.215)			
Public non-Union	0.344 (0.164)**	0.462 (0.17)***	0.039 (0.207)			
changepublicD				0.32 (0.131)**	0.335 (0.147)**	0.031 (0.189)
Private Union	0.125 (0.12)	0.107 (0.125)	0.144 (0.132)			
Casual	-.255 (0.135)*	-.296 (0.142)**	-.170 (0.143)	-.304 (0.131)**	-.342 (0.134)**	-.215 (0.141)
Self employed	-.147 (0.186)	-.155 (0.192)	-.181 (0.221)	-.177 (0.183)	-.182 (0.187)	-.204 (0.219)
Const.	-.107 (0.162)	-.019 (0.147)	-.222 (0.18)	-.084 (0.162)	-.002 (0.151)	-.202 (0.18)
Obs.	628	585	570	628	585	570
R ²	0.084	0.076	0.062	0.079	0.072	0.059

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Standard errors are in parentheses and adjusted for intra-cluster correlation. The dependent variable is the log of hourly net income. Year dummies included but not reported.

Table 14: MALE AND FEMALE REGRESSIONS

	OLS Female	OLS Male	Fixed Effects Female	Fixed Effects Male
	(1)	(2)	(3)	(4)
Age	0.053 (0.016)***	0.069 (0.016)***	0.073 (0.049)	0.138 (0.038)***
Age ²	-0.0005 (0.0002)***	-0.0007 (0.0002)***	-0.0005 (0.0006)	-0.002 (0.0005)***
Incomplete prim educ	0.142 (0.083)*	0.465 (0.115)***		
Complete prim educ	0.482 (0.143)***	0.525 (0.122)***		
Incomplete sec educ	0.708 (0.115)***	0.775 (0.13)***		
Complete sec educ	0.994 (0.169)***	1.078 (0.15)***		
Tertiary educ	1.567 (0.171)***	1.399 (0.176)***		
Married	0.007 (0.062)	0.069 (0.064)		
Public Union	0.971 (0.137)***	0.831 (0.108)***	0.631 (0.306)**	0.676 (0.211)***
Public non-Union	0.75 (0.135)***	0.557 (0.102)***	0.523 (0.267)*	0.419 (0.178)**
Private Union	0.411 (0.107)***	0.536 (0.101)***	0.104 (0.162)	0.16 (0.111)
Casual Employment	-0.244 (0.072)***	-0.353 (0.101)***	-0.328 (0.167)**	-0.083 (0.163)
Self employment	-0.528 (0.111)***	-0.339 (0.147)**	-0.047 (0.304)	-0.206 (0.206)
Const.	-0.301 (0.329)	-0.431 (0.331)	-0.372 (0.962)	-0.792 (0.727)
Obs.	1300	1306	1337	1338
R ²	0.42	0.328	0.093	0.077

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Standard errors are in parentheses and adjusted for intra-cluster correlation. The dependent variable is the log of hourly net income. Year Dummies included but not reported. Reference employment category is private non-union regular employment.

Table 15: IPW AND OLS EARNINGS FUNCTIONS

	IPW	Stayers
	(1)	(2)
Age	0.073 (0.016)***	0.067 (0.019)***
Age ²	-0.0008 (0.0002)***	-0.0007 (0.0002)***
Incomplete prim educ	0.306 (0.078)***	0.374 (0.096)***
Complete prim educ	0.56 (0.105)***	0.611 (0.123)***
Incomplete sec educ	0.677 (0.08)***	0.811 (0.098)***
Complete sec educ	1.015 (0.102)***	1.196 (0.125)***
Tertiary educ	1.474 (0.158)***	1.471 (0.159)***
Married	0.092 (0.057)	0.043 (0.066)
Male	0.241 (0.05)***	0.285 (0.061)***
Public Union	0.881 (0.103)***	0.927 (0.114)***
Public non-Union	0.732 (0.091)***	0.743 (0.111)***
Private Union	0.523 (0.064)***	0.473 (0.087)***
Casual Employment	-0.336 (0.07)***	-0.224 (0.088)**
Self employment	-0.439 (0.114)***	-0.138 (0.094)
Const.	-0.674 (0.299)**	-0.624 (0.388)
Obs.	2229	1131
R^2	0.372	0.353

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Standard errors are in parentheses and adjusted for intra-cluster correlation. The dependent variable is the log of hourly net income. Year Dummies included but not reported. Reference employment category is private non-union regular employment. The first column corrects for attrition in OLS using Inverse Probability Weighting. The second column shows the OLS results for the sub-sample of earners who are observed in more than one wave.

Table 16: ATTRITION CORRECTED DIFFERENCED EARNINGS FUNCTIONS

	First Diff	Second Diff	FirstDiff 93 98	FirstDiff 98 04
	(1)	(2)	(3)	(4)
Age	0.055 (0.029)*	-0.434 (0.318)	0.05 (0.062)	-0.010 (0.044)
Age ²	-0.0007 (0.0004)**	0.004 (0.003)	-0.0009 (0.0005)*	0.00005 (0.0005)
Public Union	0.438 (0.158)***	0.473 (0.343)	0.392 (0.181)**	0.641 (0.242)***
Public non-Union	0.294 (0.169)*	1.067 (0.437)**	0.133 (0.207)	0.778 (0.257)***
Private Union	0.097 (0.13)	0.028 (0.235)	0.088 (0.138)	0.251 (0.166)
Casual	-0.265 (0.133)**	-0.281 (0.314)	-0.138 (0.211)	-0.345 (0.131)***
Self employed	-0.176 (0.187)	-0.421 (0.243)*	-0.147 (0.228)	-0.121 (0.319)
1998 Year Dummy	0.619 (0.158)***			
lambda9898	-0.205 (0.106)*			
lambda0404	0.328 (0.092)***			
lambda93930404		-0.390 (0.6)		
Obs.	594	110	413	215
R^2	0.098	0.15	0.03	0.088

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Standard errors are in parentheses and adjusted for intra-cluster correlation. The dependent variable is the log of hourly net income. All variables are dummies except age variables. Reference employment category is private non-union regular employment.

Table 17: INDIVIDUAL ATTRITION PROBIT REGRESSIONS

	Attrition 93-98	Attrition 98-2004	Attrition 93-2004
	(1)	(2)	(3)
Age 13-19	0.116 (0.059)*	0.324 (0.059)***	0.114 (0.052)**
Age 22-30	0.305 (0.068)***	0.282 (0.067)***	0.07 (0.057)
Age 31-50	0.141 (0.08)*	0.231 (0.072)***	0.014 (0.069)
Age 51-64	0.104 (0.102)	0.172 (0.113)	-0.099 (0.092)
Age 65+	-0.178 (0.1)*	0.057 (0.099)	-0.290 (0.09)***
Male	-0.061 (0.032)*	0.044 (0.032)	0.025 (0.03)
Married	0.203 (0.066)***	0.126 (0.071)*	0.08 (0.055)
Resident Dummy	-0.177 (0.069)**	-0.255 (0.062)***	-0.088 (0.063)
Employed	0.089 (0.083)	-0.032 (0.085)	0.05 (0.076)
Public Union	-0.322 (0.238)	0.077 (0.157)	-0.326 (0.199)
Public non-Union	-0.225 (0.189)	0.336 (0.169)**	0.003 (0.157)
Private Union	-0.334 (0.156)**	0.137 (0.141)	-0.077 (0.131)
Private non-Union	-0.013 (0.111)	0.17 (0.114)	0.01 (0.102)
Casual Employment	0.003 (0.14)	0.194 (0.134)	0.152 (0.132)
Self employment	-0.187 (0.13)	0.067 (0.137)	-0.045 (0.106)
Incomplete prim educ	-0.051 (0.046)	-0.107 (0.046)**	0.135 (0.039)***
Complete prim educ	-0.100 (0.078)	-0.027 (0.073)	0.191 (0.067)***
Incomplete sec educ	0.012 (0.062)	0.006 (0.064)	0.159 (0.055)***
Complete sec educ	-0.037 (0.09)	0.021 (0.084)	0.194 (0.082)**
Tertiary educ	-0.036 (0.189)	0.068 (0.178)	0.308 (0.163)*

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Marginal effects reported. The dependent variable is a dummy, with a 1 indicating an individual was not observed in the following wave. Regression continued in Table 18.

Table 18: INDIVIDUAL ATTRITION PROBIT REGRESSIONS CONTINUED
Attrition 93-98 Attrition 98-2004 Attrition 93-2004

	(1)	(2)	(3)
Core hh member	-0.478 (0.072)***	-0.563 (0.079)***	-0.406 (0.066)***
Log(residents)	-0.301 (0.077)***	-0.456 (0.059)***	-0.399 (0.034)***
Log(non-residents)	-0.019 (0.073)	-0.075 (0.059)	-0.044 (0.031)
Age of hh Head	0.003 (0.003)		0.0005 (0.001)
HH head is female	0.011 (0.079)	0.025 (0.07)	-0.018 (0.035)
Max educ level in hh	-0.002 (0.015)	-0.006 (0.018)	-0.031 (0.007)***
HH has flush toilet	0.323 (0.221)	-0.234 (0.238)	0.097 (0.118)
2nd income quintile	0.116 (0.108)	-0.097 (0.094)	0.142 (0.044)***
3rd income quintile	0.082 (0.116)	-0.067 (0.102)	0.108 (0.047)**
4th income quintile	0.095 (0.121)	-0.046 (0.105)	0.117 (0.052)**
5th income quintile	0.191 (0.136)	-0.090 (0.132)	0.193 (0.065)***
Community violence Dummy	0.102 (0.127)		0.199 (0.056)***
Rural area in 93	0.609 (0.266)**	0.356 (0.472)	0.003 (0.178)
Tarred road Dummy	0.189 (0.246)		-0.211 (0.176)
Old KwaZulu area	-0.635 (0.238)***	0.029 (0.608)	-0.547 (0.255)**
HH owns its house	0.002 (0.153)		-0.073 (0.064)
Obs.	8258	7654	8258

Notes: *, **, *** denote significance at 10%, 5% and 1% levels. Marginal effects reported. The dependent variable is a dummy, with a 1 indicating an individual was not observed in the following wave. The omitted category is a single, female non-resident, non-labour force participant, with zero years of education and who is not a core member in a household in the lowest income quintile, not owning the house the household lives in, without a flush toilet in a community with no reported violence, no tarred roads and that was located in an urban area in the old Natal province in the original 1993 survey.

References

- Abowd, John, Francis Kramarz, and David Margolis**, “High Wage Workers and High Wage Firms,” *Econometrica*, 1999, *67* (2), 251–333.
- Badaoui, Eliane, Eric Strobl, and Frank Walsh**, “Is There An Informal Sector Wage Penalty? Evidence from South Africa,” *Economic Development and Cultural Change*, 2008, *56*, 683–710.
- Bennell, Paul**, “Rates of return to education: Does the conventional pattern prevail in sub-Saharan Africa?,” *World Development*, January 1996, *24* (1), 183–199.
- Bhorat, Haroon, Carlene van der Westhuizen, and Sumayya Goga**, “Analysing Wage Formation in the South African Labour Market: The Role of Bargaining Councils,” 2007. Paper presented at 2008 DPRU conference.
- Botelho, Fernando and Vladimir Ponczek**, “Segmentation in the Brazilian Labor Market,” *Economic Development and Cultural Change*, 2011, *59* (2), 437 – 463.
- Butcher, Kristin F. and Cecilia E. Rouse**, “Wage effects of unions and industrial councils in South Africa,” *Industrial and Labor Relations Review*, January 2001, *54* (2), 349–374.
- Cameron, A. Colin and Pravin K. Trivedi**, *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press, 2005.
- Chen, Marta**, “Rethinking the Informal economy: linkages with the formal economy and the formal regulatory environment,” in Basudeb Guha-Khasnobis, Ravi Kanbur, and Elinor Ostrom, eds., *Linking the Formal and Informal Economy: Concepts and Policies*, Oxford University Press 2006.
- Cichello, Paul L., Gary S. Fields, and Murray Leibbrandt**, “Earnings and Employment Dynamics for Africans in Post-apartheid South Africa: A Panel Study of KwaZulu-Natal,” *Journal of African Economies*, June 2005, *14* (2), 143–190.
- Deaton, Angus**, *The Analysis of Household Surveys – A Microeconomic Approach to Development Policy*, Baltimore: Johns Hopkins University Press, 1997.
- Falco, Paolo, Andrew Kerr, Neil Rankin, Justin Sandefur, and Francis Teal**, “The returns to formality and informality in urban Africa,” *Journal of Labour Economics*, 2011, *18*, Supplement 1, S23–S31.
- Fields, Gary S.**, “Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs,” *Journal of Development Economics*, 1975, *2* (2), 165–187.

- , “A guide to multisector labor market models,” Technical Report 2005.
- Griliches, Zvi and Jerry A. Hausman**, “Errors in variables in panel data,” *Journal of Econometrics*, February 1986, *31* (1), 93–118.
- Günther, Isabel and Andrey Launov**, “Informal employment in developing countries. Opportunity or last resort?,” *Journal of Development Economics*, January 2012, *97* (1), 88–98.
- Harris, John and Michael. Todaro**, “Migration, Unemployment and Development: A Two-Sector Analysis,” *American Economic Review*, 1970, *60* (1), 126–142.
- Heckman, James J.**, “Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture,” *Journal of Political Economy*, August 2001, *109* (4), 673–748.
- Heckman, James J and Guilherme Sedlacek**, “Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market,” *Journal of Political Economy*, December 1985, *93* (6), 1077–1125.
- Heckman, James J. and V.J. Hotz**, “An investigation of the labor market earnings of Panamanian males: Evaluating the sources of inequality,” *Journal of Human Resources*, 1986, *23*, 507–542.
- Heintz, James and Dori Posel**, “Revisiting informal employment and segmentation in the South African labour market,” *South African Journal of Economics*, 2008, *76* (1), 26–44.
- Hussmans, Ralph**, “Measuring the informal economy: From employment in the informal sector to informal employment. ILO Working Paper 53,” Technical Report, ILO, Washington, DC 2004.
- International Monetary Fund**, “International Financial Statistics September 2009, Series 199..AF.ZF,” 2009.
- Keswell, Malcolm and Laura Poswell**, “Returns To Education In South Africa: A Retrospective Sensitivity Analysis Of The Available Evidence,” *South African Journal of Economics*, 2004, *72* (4), 834–860.
- Kingdon, Geeta and John Knight**, “Unemployment in South Africa: The Nature of the Beast,” *World Development*, March 2004, *32* (3), 391–408.
- Magruder, Jeremy**, “High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa,” 2009. University of California, Berkeley, Working Paper.
- Maloney, William**, “Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico,” *The World Bank Economic Review*, 1999, *13* (2), 275–302.

- , “Informality Revisited,” *World Development*, July 2004, *32* (7), 1159–1178.
- May, Julian, Jorge Agüero, Michael R. Carter, and Ian M. Timaeus**, “The KwaZulu-Natal Income Dynamics Study (KIDS) 3rd wave: methods, first findings and an agenda for future research.,” *Development Southern Africa*, 2007, *24*, 629 – 648.
- Moll, Peter**, “Compulsory Centralization of Collective Bargaining in South Africa,” *American Economic Review*, 1996, *86*, 326–329.
- Pratap, Sangeeta and Erwan Quintin**, “Are labor markets segmented in developing countries? A semiparametric approach,” *European Economic Review*, October 2006, *50* (7), 1817–1841.
- Psacharopoulos, George**, “Returns to investment in education: A global update,” *World Development*, September 1994, *22* (9), 1325–1343.
- Roy, A.**, “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, 1951, *3* (2), 135–146.
- Wooldridge, Jeffrey M.**, *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA.: MIT Press, 2002.