Southern Africa Labour and Development Research Unit

Earnings volatility in South Africa

by
Vimal Ranchhod

About the Author(s) and Acknowledgments

Vimal Ranchhod - Chief Research Officer, SALDRU, Dept. of Economics, University of Cape Town. Email:
vimal.ranchhod@gmail.com

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R15.00 each plus vat and postage charges.

Orders may be directed to:
The Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,
Tel: (021) 650 5696, Fax: (021) 650 5697, Email: brenda.adams@uct.ac.za

Earnings volatility in South Africa

V. Ranchhod[1]

Abstract

How much volatility is there in earnings in South Africa? The South African labour market has been shown to be a key determinant of welfare, both in terms of poverty and inequality. These are a function of both the high levels of unemployment as well as the wage distribution, conditional on being employed. One aspect of welfare that derives from the labour market, which has been relatively understudied to date, is the amount of volatility in earnings that various groups of South Africans experience over time. This has implications directly for welfare, as well as for inequality. We make use of the first three waves of data from the National Income Dynamics Study to describe the amount of earnings volatility experienced by different demographic groups. We then make use of a regression model to estimate the partial correlation between the various characteristics that we use and earnings volatility. Our main findings are that earnings volatility is high over a four year interval. The mean within-person standard deviation in earnings across the three waves lies between 50% and 66% of the mean earnings depending on the time period, and the mean within-person coefficient of variation in earnings is 0.641.

---

[1] Chief Research Officer, SALDRU, Dept. of Economics, Univ. Of Cape Town. Email: vimal.ranchhod@gmail.com

1. Introduction

The South African labour market is one of the central mechanisms through which individual and household welfare is determined. Household poverty levels are strongly affected by whether a household has at least one member who has formal sector employment (Leibbrandt, Bhorat and Woolard, 2001). In addition, South Africa has one of the most unequal income distributions in the world, and decompositions reveal that most of this is due to wage inequality. A substantial fraction of the wage inequality, in turn, is determined by the people who earn zero wages, i.e. the unemployed. In this paper we aim to further our understanding of the welfare dynamics that are generated by the labour market, by focussing on the volatility of earnings in the South African labour market.

What do we mean by "earnings volatility"? The first term, "earnings", refers to that subset of income that is derived from one's supply of labour in the labour market. Put differently, it is the wages that a person or group earns in return for the time that they spend working. The word "volatility", in our context, refers to how quickly and sharply the object under observation is likely to change. Putting these terms together, a workable definition for "earnings volatility" is that it is a concept that describes the frequency and magnitudes with which wages change over time in the South African labour market. Greater volatility in earnings would arise if wages start to change more rapidly, or if the magnitude of wage changes were to increase, or both. Since the movement from employment to unemployment, or vice versa, entail large effective changes in wage levels, a large fraction of earnings volatility is likely to be accounted for by transitions into and out of employment.

It is also important to differentiate between the volatility experienced by an individual and that experienced by a group of individuals. Our study is concerned with the average volatility of individuals within a group. For example, we might be interested in the earnings volatility that young African women experience over a four year period. The object of interest is at the group level, but the group is comprised of many individuals. In this example, we will make use of individual level data to estimate first the volatility of earnings for each young African woman in our sample, and then calculate the average volatility within the group. This is potentially different from simply estimating the cross-sectional variation in earnings amongst the group of young African women.

Understanding earnings volatility is potentially important from a social welfare perspective. Uncertainty over one's future stream of income, which is related to earnings volatility, is likely to negatively affect people's utility. From a theoretical perspective, this follows from the conventional assumption that people are risk averse, which is captured in formal economic models by the concavity of the utility function. Practically, we could get volatility in earnings that are
 anticipated or unanticipated. The former might affect welfare as financial markets may be imperfect and credit constraints might be binding. The latter affects welfare as insurance markets may be imperfect. The possibility of income shocks might result in hoarding of resources for precautionary motives, or the need to incur costly debt for unanticipated costs.

An additional dimension where volatility could have an effect is on measures of inequality. Suppose that there exists an economy with only two people, who live for only two time periods. Consider the following three scenarios:

- In scenario 1, each person receives wages of R100 in each time period.
- In scenario 2, one person receives a wage of R200 in the first time period, while the other person is unemployed. In the second time period, the first person becomes unemployed and the second gets a wage of R200.
- In the third scenario, one person receives a wage of R200 in each time period while the other remains unemployed in both time periods.

In comparing the three scenarios, it should be clear that the first scenario entails a world of perfect equality, both from a cross-sectional and dynamic perspective. The second scenario has extreme inequality in either time period, but is egalitarian from a dynamic perspective. Here, the volatility from the labour market reduces the dynamic levels of inequality. In the third scenario, we also have extreme inequality in each time period, but the labour market is stable. In this case, intuitively at least, it would seem that the long run inequality is even more pronounced than the contemporaneous inequality in any time period. The general point of this illustrative exercise is that earnings volatility is intrinsically related to economic mobility, and economic mobility is likely to reduce the level of inequality measured over a longer time horizon, relative to cross-sectional measures of inequality.

The purpose of this paper is to describe the levels of earnings volatility in the South African labour market, using the first three waves of the National Income Dynamics Study (NIDS) from 2008 to 2012. Related literature in the national South African context is virtually non-existent, most likely due to the lack of availability of nationally representative longitudinal datasets. Some region specific literature has been generated, such as the work by Cichello, Fields, and Leibbrandt (2005). The authors make use of the KIDS data, and examine the probability of experiencing various poverty transitions as a function of employment changes over 1993–98. One of their overall findings is that volatility in employment and earning status is a major determinant poverty.

In the US, a much longer established literature has emerged based mostly on the Panel Study of Income Dynamics (PSID), although other longitudinal datasets have also been used. The literature is relatively advanced, primarily due to the long run nature of the PSID, which has been in existence since 1968. This essentially allows researchers to observe respondents' earnings over much of their working lives for respondents who were first interviewed in the earlier waves of the PSID. For example, the seminal article in this literature, by Gottschalk and Moffitt (1994), has already been cited 749 times. In that paper, the authors used PSID data from 1970 – 1987, and found evidence that suggested that the increase in men's earnings inequality arose partially due to an increase in transitory earnings variation.

Extensions to their paper used different datasets, different ways of measuring volatility and different ways of decomposing volatility into permanent and transitory components. A recent paper by Shin and Solon (2011) suggests that there remains considerable debate about trends in earnings volatility in the US, and partly this is due to the different ways in

which it has been measured. While much of the literature has been motivated by trying to understand the link between earnings volatility and income inequality, some researchers have focussed on the importance of volatility as a component of welfare in its own right.

In his popular book, titled *The Great Risk Shift: The Assault on American Jobs, Families, Health Care and Retirement And How You Can Fight Back*, Hacker (2006) makes use of PSID data and claims that "... the volatility of family incomes has gone up – way, way up ... In fact, over the past generation the economic *instability* of American families has actually risen much faster than economic *inequality* ...".[2] He then goes on to argue that this has been caused by a sustained transfer of risk from corporations to the middle and working classes, and that this is likely to generate economic inefficiencies.

In comparison, our objectives in this paper are modest. First, we do not (yet) have the dataset available to analyse long run trends in volatility, and second, the US literature has been evolving for about two decades already. The corresponding SA literature is quite undeveloped, and this paper represents an early attempt at getting this line of research going. As such, our analyses here are descriptive and relatively simple. We first estimate the transition rates between employment and unemployment for different groups of individuals, defined by age, race, gender, education and geographical location. This is initially done only for the employment/unemployment margin and for the short 4 year period over which we have observed our respondents. Next, we consider the observed wages of these same respondents, where the unemployed earn a zero wage. We then compute the average volatility of earnings amongst respondents in a group, over the same period of observation. Our final analysis involves estimating a regression model of earnings volatility with the various demographic characteristics as our explanatory variables. This allows us to estimate how volatility correlates with some characteristics, such as levels of educational attainment, while controlling for all of the other characteristics. Education, in particular, is an important variable because it is one of the few variables that we consider that can be influenced by policy decisions.

Our overall findings are that South Africans experience substantial amounts of earnings volatility, especially given the short time period that we consider. The mean within-person standard deviation in earnings across the three waves lies between 50% and 66% of the mean earnings depending on the time period that we use to calculate the mean earnings, and the mean within-person coefficient of variation in earnings is 0.641.

The remainder of this paper is structured as follows. In section 2 we describe the methods that we use. We describe in detail the analyses that we undertake as well as any relevant assumptions and their potential implications. Section 3 presents the data and corresponding sample sizes used in our analyses. Section 4 presents and discusses the results, and we conclude in Section 5.

---

[2] Hacker, 2006, p.2

2. Methods

Throughout this paper, we consider the levels of earnings volatility as observed amongst different groups of people. The groups are defined by various demographic characteristics.

Our analysis involves three different calculations. First, we estimate the observed likelihood that a particular group either finds employment or loses employment. These are essentially two period transition matrices. For example, suppose that the group that we are interested in has some tertiary level educational qualification. Within this group, some people in wave 1 are employed, while some are not. For the subset of this group that are employed, some fraction will still be employed in both waves 2 and 3, some will be employed in wave 2 but lose their employment in wave 3, some will be employed in wave 2 but then fall back into unemployment in wave 3, and some will be unemployed in both waves 2 and 3. An identical categorization can be used to classify the subsequent employment status of the unemployed subset of people in wave 1 who have some tertiary level qualification.

How would this allow us to generate a measure of volatility? South Africa has high levels of unemployment, and we assume that losing or finding a job is likely to have a large impact on one's earnings relative to wage variation within the group of stably employed people. We thus expect that changes in a person's employment status would yield a reasonable proxy measure of earnings volatility, as these changes involve switching between a positive wage and a zero wage. Thus, at the group level, we summarize the percentage in the group that experience 0, 1 or 2 changes in their employment status over the three waves of NIDS. Note that, subject to data availability, this system of classification is mutually exclusive and exhaustive, regardless of the group under consideration. Earnings volatility is likely to be highest amongst the sub-population where we observe one or two changes in their employment status between waves 1 and 3, and volatility is likely to be lowest amongst people who experience no change in their employment status over the period of observation.[3]

The second component of our analysis is to calculate the real earnings of each person in each of the waves, and then estimate the within-person standard deviation in wages. As stated previously, respondents who were not employed were assigned a wage of zero. The real wages were calculated using the CPI from StatsSA and calibrated to July 2012.[4] The within person standard deviation is a measure of the spread of an individual's wages over time, in relation to that person's average wage over the three points in time. Given that our working definition of earnings volatility emphasized the speed and magnitude of changes in earnings, the standard deviation is a useful statistical measure as it will increase if either the speed or magnitude of changes were to increase.

At the same time, one of the properties, and thus limitations, of standard deviations is that they will tend to be larger for people who earn high wages relative to those who do not.

---

[3] Indeed, amongst people who are unemployed in each of the three waves, earnings volatility must be zero over the three waves since they would not be observed to be receiving any wage income over the duration of the study.

[4] We multiplied the wave 1 wages by 97.8/81.2, which is the July 2012 CPI value divided by the July 2008 CPI value. Similarly, we inflated the wave 2 wages by 97.8/88.6, where 88.6 is the July 2010 CPI value.

This could result in potentially misleading comparisons across groups. For example, people with higher levels of education might earn substantially more than those who have not finished matric, and the employment prospects of the latter group may indeed be more unstable, yet we could still find that the mean of the standard deviation in earnings is higher amongst the better educated group. For this reason, we additionally perform this same analysis using the coefficient of variation instead of the standard deviation. This is a statistic that normalizes the within-person standard deviation of earnings by the within-person mean earnings, and thus allows a better comparison across groups of people who have different mean levels of earnings. We present the mean of the (within-person) standard deviation in earnings, as well as the mean of the (within-person) coefficient of variation, for each of our groups.

The final component of our analysis is to estimate regression models of earnings volatility on a set of dummy variables that identifies our various groups. We estimate two such regression models, where the dependent variables are the within-person standard deviation and within-person coefficient of variation respectively. Both measures of volatility are useful, as the former measures volatility in terms of rands, while the latter is a unitless measure that does not have the 'scale' problem discussed in the preceding paragraph.

3. Data

The dataset that we use for this paper is the balanced component of the first three waves of NIDS. We restrict our sample to observations that appear in all three waves, as well as observations that have a valid response for each of the variables that we use. NIDS is a large, nationally representative study, which interviews people at approximately biannual intervals.

As with any longitudinal study, one is generally concerned with how much attrition there is over time. In Table 1 below, we show how our sample changes as we impose increasingly stringent sample restrictions. Starting with wave 1 from 2008, and including respondents to either the Adult or Proxy questionnaire, we have a maximum possible sample of 18621. This number decreases to 17231 observations that we manage to merge across the three waves, or an attrition rate of 7.6%. Once we impose our age restriction to include only the working aged population in wave 1, our sample decreases further to 15182 respondents, with a net attrition rate of 18.5%.

Table 1: Sample sizes and attrition

Description	N	% reduction
Wave 1 adults	16871	
Wave 1 Proxy	1750	
Maximum sample size	**18621**	
Merged across all three waves	17213	0.076
Age restriction (16 - 64 in wave 1)	15182	0.185
Valid panel weight	9970	0.465
Final balanced panel	**9016**	**0.516**

Weights

An important and challenging process that was necessary was to generate an appropriate set of weights for the balanced panel. NIDS did not release such panel weights for the balanced panel with the official release, and hence the onus rests with researchers to find an appropriate set of weights. The objective of using weights is to re-scale the data so that the subsequent weighted sample represents that of the population being studied. In the wave 1 data, there is thus a set of design weights that correct for differences in the probability of being selected into the sample. These weights are then adjusted for non-response in wave 1.

To calculate panel weights, we estimated two sets of probit models. The first was for the probability of successfully re-interviewing a wave 1 respondent in wave 2, and the second was for the probability that a respondent who was interviewed in both wave 1 and wave 2 was also successfully interviewed in wave 3. Each probit included race dummies, a male dummy, a highly flexible set of age variables, an educational attainment variable, a marital status dummy, province dummies, and a set of dummies for the type of area that the household was located in, in the previous wave. From each probit, we predict the probability of successfully re-interviewing the individual in the subsequent wave. The balanced panel weights are then computed as the original design weights adjusted for non-response, multiplied by the inverse of the predicted probability of successfully being re-interviewed in wave two, which is then further multiplied by the inverse of the predicted probability of successfully being re-interviewed in wave 3, conditional on having successfully been interviewed in both waves 1 and 2.

An unfortunate outcome of this process is that we have several respondents for whom we do not have a valid panel weight. In this case, these individuals are subsequently removed from the sample used for our analyses. There are many reasons why we do not have a higher number of valid panel weights. First, if the person has a missing value for even one of the covariates used in either probit model, we will not obtain a valid panel weight. This is particularly true for anyone with a Proxy interview in any wave. Second, the estimation process used to estimate the probit coefficients may fail to converge for certain sub-populations, which will also result in these people being excluded from the analysis. Third, we trimmed the weights to exclude outliers. While there is no clear best practise in this situation, any choice has advantages and disadvantages. The presence of outliers in the weights indicates that a particular respondent who survived across all three waves was highly unusual, relative to their peers who had the same demographic characteristics. Including such respondents with very high weights would then imply that we are willing to assume that this unlikely person is nonetheless representative of the group of people who

had similar characteristics in wave 1. We felt that this was implausible. In addition, not trimming our weights would have affected our analysis in an economically meaningful and potentially misleading way.

The net result of using these weights on our sample size is substantial. From Table 1, we see that our sample drops from 15182 to 9970 when we require that a person has a valid weight. This is a loss of about one third of the remaining sample.[5] Our sample is further reduced by 954 observations which have either missing employment status or missing earnings information, in at least one of the three waves. Thus, our balanced panel sample size used in this analysis is 9016, which represents a 51.6% loss in sample relative to a hypothetical maximum that would have included all adults in wave 1.

Variables
We make use of a few variables for our analysis. The variables that define our demographic groups, and how they were coded are:

- Age: We construct three age groups based on their age in Wave 1, namely youth aged 16 to 29, prime aged adults aged 30 to 49 and older adults aged 50 to 64.
- Race: For some of our analysis, we focus only on the African sub-sample. There are too few White and Indian respondents for meaningful racial comparisons, and we did not expect to find much of interest in comparing differences in earnings volatility between the African and Coloured sub-samples.
- Gender: We compare the volatility amongst female and male respondents.
- Education: We compare the volatility observed by different levels of educational attainment as measured in wave 1. To implement this component of the analysis, we needed to classify different levels of education into a small number of categories. We decided to create three groups, one for people who have not completed secondary school, one for people who only have a matric qualification, and one for people who have any form of a post-matric qualification. This categorization was used as it seems to conform with different levels of signals of human capital in the labour market. In all likelihood, employers would also differentiate between a four year university degree and a 6 month diploma, but our group sizes become too small for such a comparison.
- Geographic location: We separated the sample into respondents residing in urban or rural areas in wave 1. Labour markets are likely to operate quite differently in urban and rural areas, and the job prospects and wage distributions are also likely to differ substantially between them. This led us to consider a comparison along these lines.

In addition to the demographic variables defined above, we use two sets of variables either as outcome variables, or to derive our outcome variables. These are:

- Employed_w1, Employed_w2, Employed_w3: These are binary variables that identify whether a person was employed in the respective wave. The variables include data

---

[5] We consider whether the balanced panel nonetheless remains a useful representation of the cross-sections in our discussion of Table 3 below.

from the proxy surveys where this was useful. Note that many different forms of employment are captured by this variable, including regular employment, casual employment or self-employment. In addition, we chose not to expand the number of labour market states to differentiate between the unemployed and the "not economically active". This decision was a pragmatic one as it aids with the analysis and exposition of results tremendously, while not affecting earnings volatility in a substantial way.

- Earnings_w1, Earnings_w2, Earnings_w3: These are the total earnings from regular employment, casual employment or self-employment, in a particular wave. They are bounded below by zero and are assigned a value of zero if the person was not employed in that wave. The earnings variables are converted to real values using the July 2012 CPI value as the baseline.

As already discussed in the methods section, we use the earnings variables to calculate our two dependent variables for our final set of analyses, namely the within-person standard deviation in earnings (across the three waves), and the within-person coefficient of variation (across the three waves).

Sample composition and descriptive statistics
In Table 2 below we show how the composition of the balanced sample compares with each cross-section, in terms of the demographic groups that we consider.

We note several interesting observations from Table 2. First, the sample size grows quite substantially with time in cross-sections, due to new members of the household. In relative terms, the sample size in the balanced panel is much smaller. In addition to not containing new household members, the balanced panel is also smaller due to attrition from the wave 1 cross-section. Nonetheless, the total sample size remains large enough to provide sufficient statistical power for the analyses that we undertake in this paper.

Of the variables that we focus on, we notice that both the age and gender distributions in the panel are reasonably close to that in wave 1, but not in wave 2 and 3. This is likely due to a mixture of demographic shifts (i.e. fertility and mortality), as well as household composition changes. In addition, the sample becomes more African with time, and the panel over-represents Africans, especially with respect to the wave 1 cross-section. This reflects that attrition is disproportionately a non-African phenomenon.

When we compare the distribution of educational attainment between each cross-section and the balanced panel, we observe that the panel over-represents high school dropouts and under-represents matriculants and people with a tertiary qualification, relative to the wave 1 cross-section. With time, education levels are increasing and so our panel becomes slightly less representative of later waves. In terms of urban/rural spread, the differences between the panel and the cross-sections are quite small and representativity does not seem to be a substantial issue along this dimension.

Table 2: Composition of sample: balanced panel and cross sections

Variable	Cross sections			Panel
	Wave 1	Wave 2	Wave 3	All_waves
Sample size	16131	20177	22186	9016
Age distribution				
% youth (16 - 29)	43.72	46.42	46.26	43.32
% prime aged (30 - 49)	37.85	36.16	36.24	38.23
% Older (50 - 64)	18.43	17.42	17.5	18.44
Race				
% African	76.54	78.99	79.28	81.08
Gender				
% male	40.78	44.7	44.72	40.26
% female	59.22	55.3	55.28	59.74
Education				
% < matric	74.62	73.91	71.78	77.04
% matric	17.12	18.15	17.8	15.62
% some tertiary	8.26	7.94	10.42	7.34
Location				
% Rural	48.11	50.8	49.74	51.4
% Urban	51.89	49.2	50.26	48.6

Notes:
1. The panel classifications are based on wave 1 data where possible.
As such, we woud expect there to be a stronger similarity between
the wave 1 cross section and the panel, than between the panel
and wave 2 or wave 3.
2. The cell proportions are unweighted for both the panel and each cross-
section.

Descriptive Statistics

In Table 3 below, we present a comparison of the mean proportion employed and mean earnings in each wave of the cross section as well as the corresponding subset of the balanced panel. All the means and proportions are weighted. In the cross-sections, we use the conventional design weights which are adjusted for non-response. In the balanced panel, we use the panel weights that we generated. Also, the age categories in the cross-sections are shifted up by two and four years in wave 2 and wave 3 respectively. We did this in order to maintain comparability with the panel age groups, which are defined based on their wave 1 ages.

The overall point that we obtain from the comparison is that the panel summary statistics yield a more optimistic picture than the cross-sections, both in terms of earnings as well as employment. There are two ways in which this could arise, and these are not mutually

exclusive. First, there is the effect of attrition in the panel. If attrition is correlated with having poorer outcomes along these dimensions, then the respondents who survive into the balanced panel will have better labour market experiences than the cross-section on the whole. Second, there is the composition effect of selective migration into households and the formation of new NIDS households. People who enter into NIDS households after wave 1 are not continuing sample members (CSMs), but are nonetheless included in the wave 2 and wave 3 cross-sections. These two processes are unlikely to be independent of each other, as people without resources are probably more likely to join households where someone is employed or well paid, if possible.

We can get a sense of the magnitude of these two processes. Consider first the difference between the wave 1 panel and wave 1 cross-section, which measures only the attrition effect. Next consider subsequent divergences between the panel and cross-section in wave 2 and wave 3, which reflects the combined effects of attrition as well as selective migration and household formation. Thus, by comparing the differences in waves 2 and 3 relative to those in wave 1, we can obtain some measure of the effect of selection in household composition on the comparability of the balanced panel relative to the cross-section.

To illustrate the point, let us focus on the proportion employed in the first line in Table 3. The difference in the estimated proportion employed between the panel data and the cross-section in wave 1 is 0.003, which is very small. This implies that at baseline, the balanced panel does represent the cross-section quite well. In wave 2, this difference increases to 0.051, which then decreases to 0.035 in wave 3. In wave 2, the proportion employed in the panel remained approximately unchanged, and almost all of the difference of 0.051 occurs as a result of a decrease in the proportion employed in the cross-sectional data. This suggests that the divergence in the proportion employed between wave 1 and wave 2 is likely due to the household composition effect. Between waves 2 and 3, the proportion employed in the cross-section and in the panel increase by 5.6 and 4.0 percentage points respectively. Thus, the improvements observed reflect an improvement in employment prospects for both CSMs as well as temporary sample members (TSMs), where the improvement was larger for the TSMs. Put differently, employment prospects improved for members of the balanced panel as well as for the people that they lived with, and the improvement observed amongst the TSMs potentially reflects both a real improvement in employment prospects as well as a change in household composition.

A similar analysis of changes in mean earnings highlights the effect of attrition a bit more sharply. In wave 1 already, there is a substantial difference in mean earnings between the overall cross-section and the subset of data that survives into the panel, with the former being 1875 and the latter being 2171 rands per month.[6] Again, when we look at wave 2, we find that the gap between the cross-section and panel means has widened, but the panel mean is quite similar to the panel mean from wave 1. The increase in the gap is driven primarily by a decrease in the wave 2 cross-sectional mean, which decreased to 1726 rands per month in real terms. This reflects a household composition effect, which is consistent with the employment differences discussed in the previous paragraph. Between wave 2 and wave 3, the mean wages in the cross-section increased by 24%, while the mean wages in the

---

[6] Note that unemployed people were assigned a wage of zero.

panel increased by about 20%. The gap between the panel and cross-section thus got smaller in percentage terms. Consistent with the employment dimension, mean wages improved for both the panel members as well as the TSMs. The latter reflects improvements in both individual TSMs' earnings as well as positive household composition effects.

When we perform the comparison within the groups that we have defined, we also observe some differences between the panel waves and the corresponding waves of cross-sections. When considering age groups, the panel yields similar employment levels but substantially higher wage rates for older adults. There is also a large divergence with time for wages amongst prime aged individuals. Within the African subgroup, the panel in wave 1 accurately reflects the proportion that are employed, but nonetheless over-states mean earnings by about 15%. As with the rest of the table, these differences become more pronounced with time. Within the educational subgroups, a similar pattern is obtained. Initially, the proportion employed is reasonably similar, but this diverges sharply with time amongst those with a matric or more than a matric. Along the earnings dimension, there are substantial differences to begin with and these increase further as time goes by. The general pattern observed in all the other subgroups is also manifest when we consider the urban and rural respondents from wave 1 separately.

Overall, what this means is that we need to be cautious about generalizing from the panel to the society at large. On the employment dimension, this seems to be less of a problem, but in terms of mean earnings, the problem is more pronounced. For our purposes, this is at least somewhat helpful, since most of the volatility is obtained from movements into and out of employment. It is also useful to observe that a substantial amount of the divergence is derived from household composition effects, and not only attrition. The purpose of the NIDS panel (or any other individual level panel for that matter) cannot be to replicate changes in household composition, but rather to measure the evolution of the circumstances of a representative group of respondents over time. While not perfect, the overall finding here is that the subsequent analyses will nonetheless provide us with useful and reasonable estimates of the earnings volatility experienced by a representative group of South Africans over the period from 2008 to 2012.

Table 3: Comparison of proportion employed and mean earnings in cross sections and panels

| Variable		Wave 1		Wave 2		Wave 3	
		X-sect	Panel	X-sect	Panel	X-sect	Panel
Overall sample	prop. Employed	0.450	0.453	0.401	0.452	0.457	0.492
	mean wages	1875.77	2171.26	1726.86	2160.38	2139.79	2595.19
Age groups							
Youth	prop. employed	0.295	0.263	0.293	0.323	0.397	0.416
	mean wages	657.46	706.53	785.41	942.09	1337.73	1433.32
Prime aged	prop. employed	0.615	0.630	0.548	0.603	0.588	0.628
	mean wages	2963.94	3001.72	2665.58	3202.16	3128.69	3712.37
Older adult	prop. employed	0.469	0.469	0.366	0.386	0.329	0.333
	mean wages	2545.66	3677.80	2157.81	2519.11	2088.68	2622.35
Race							
African	prop. employed	0.414	0.415	0.369	0.418	0.432	0.469
	mean wages	1124.16	1294.83	1185.17	1500.90	1584.69	1821.91
Gender							
Male	prop. employed	0.551	0.564	0.485	0.548	0.549	0.591
	mean wages	2901.88	3180.06	2414.62	2922.70	2955.44	3520.20
Female	prop. employed	0.369	0.357	0.337	0.369	0.387	0.406
	mean wages	1182.33	1300.57	1189.33	1502.41	1510.63	1796.80
Education							
< matric	prop. employed	0.376	0.373	0.347	0.363	0.390	0.409
	mean wages	687.949	823.207	896.096	927.562	980.121	1180.49
matric	prop. employed	0.529	0.553	0.384	0.572	0.495	0.628
	mean wages	2258.60	3026.12	1675.10	3089.86	2255.67	3610.54
some tertiary	prop. employed	0.768	0.778	0.713	0.798	0.691	0.772
	mean wages	8512.98	9041.5	6909.43	8190.65	6855.85	9602.84
Location							
rural	prop. employed	0.341	0.324	0.283	0.325	0.341	0.382
	mean wages	602.001	714.21	729.900	891.618	1060.74	1188.03
urban	prop. employed	0.514	0.525	0.478	0.522	0.528	0.553
	mean wages	2541.37	2984.24	2366.594	2868.292	2785.367	3380.325

Notes:

1. Since the age categories are shifting with time, and the panel
categories are based on age in wave 1, we increase the lower
and upper age cutoffs for each age category by an additional
two and four years for the wave 2 and wave 3 cross sections respectively.
For example, youth in wave 2 are aged 18 - 31, while youth in wave 3
are aged 20 - 33 etc.
2. The urban/rural and education categories in the panel are also based on
wave 1 data, but these are time varying characteristics. We thus expect that
the panel will differ somewhat from the wave 2 and wave 3 cross-sections.
3. All cross-sectional means and % are weight by the design weights.
4. All panel means and % are weighed by the panel weights.
5. Unemployed respondents are assigned a zero wage.

4. Results

Transition rates
We next consider the likelihood of various employment trajectories across the three waves
of the panel, for our various groups. We classified respondents as either employed or
unemployed, where unemployed means "not employed", and includes the 'not
economically active' and 'discouraged' job seekers. The proportions are all obtained using
the balanced panel weights, and the results are provided graphically below.

Figure 1:
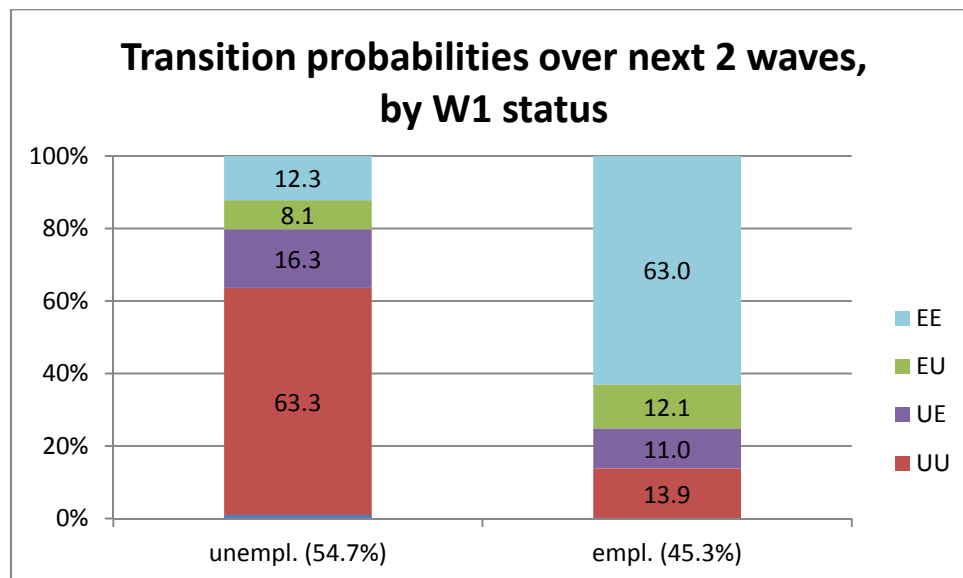
One way to interpret the graphs is to think about the red block as stable unemployment, the
purple and green blocks as unstable employment, and the blue block as stable employment.
This is because the red represents unemployment in both waves 2 and 3, the blue

represents employment in both waves 2 and 3, while the purple and green both represent a mixture of employment and unemployment over waves 2 and 3 combined. From a welfare perspective, we would ideally like to see large blue blocks and tiny red blocks in our graphs.

The graph of transitions for the full panel sample (see Fig. 1) suggests a fair amount of persistence in one's initial labour market state. Almost 2/3 of the respondents either maintain their employment, or remain in unemployment, for each of the subsequent two waves. For those who were unemployed in wave 1, this is represented by the 63.3% of people shown with the solid red block, while for those employed in wave 1 this is represented by the 63% of people shown with the solid blue block.

Approximately one quarter of the unemployed find unstable employment, while about 23% of the employed in wave 1 fall into unstable employment. Of concern is that only about 1 in 8 unemployed people transition into stable employment, while about 1 in 7 employed people experience stable unemployment in waves 2 and 3.[7]

Figure 2:

Transition probabilities by age group over W2 & W3 status: By W1 status
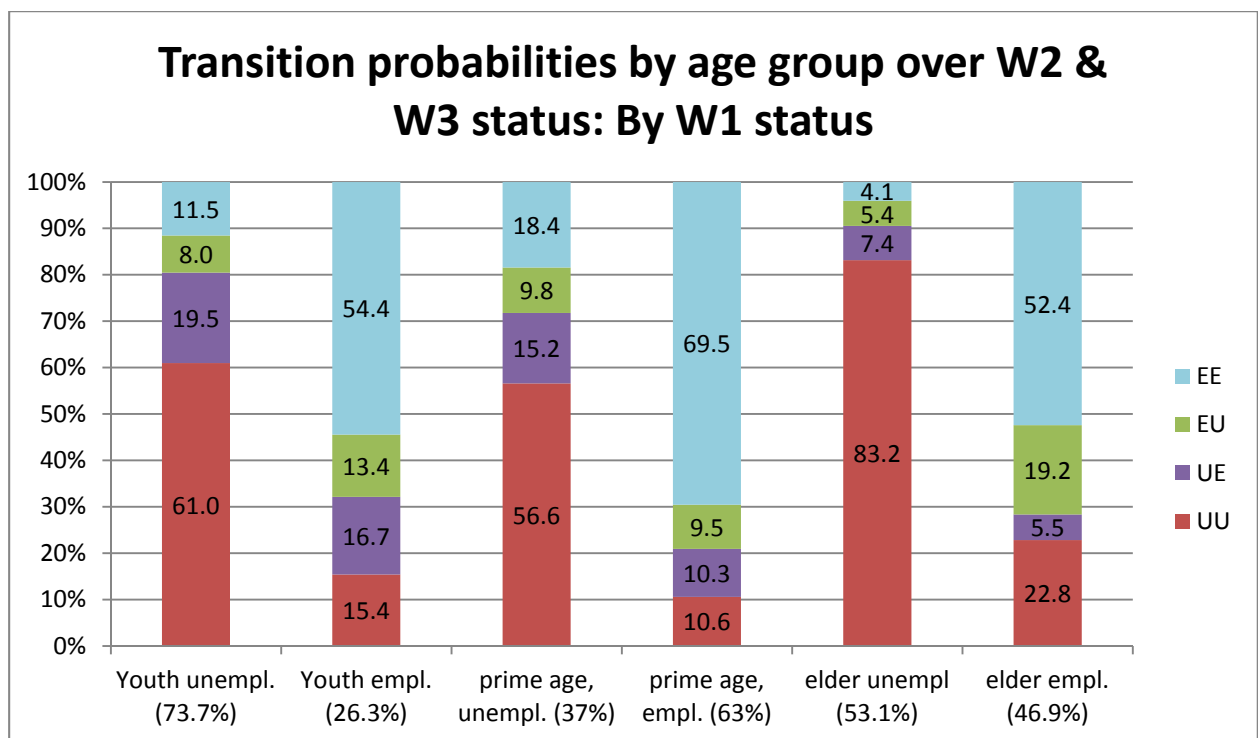
In a broad sense, the age decomposition depicted in Fig. 2 shows the pattern of life cycle employment. Stability is high amongst youth, who are potentially still studying and also have a difficult time entering the labour market. The situation remains challenging for prime aged adults in absolute terms, but is considerably better than it is for youth. As people get older and drift into retirement, the chances of finding stable employment decreases substantially.

---

[7] The word 'stable' is used loosely. We cannot say anything about spells of employment or unemployment during the approximately two year intervals between waves.

Thus, stable unemployment starts out high for youth, decreases to about 56% for unemployed prime aged adults, and increases to 83.2% for older adults. A mirror of this pattern is seen amongst the employed subsets of these age groups, when considering the likelihood of stable employment instead.

When we consider potential differences by gender, as depicted in Fig. 3 below, we observe strong gender differentials. These are consistent with other existing literature on the SA labour market. Relative to men, women are less likely to exit from unemployment, and if they do, are less likely to find stable employment. Similarly, women are more likely to lose employment, and when they do, they are more likely to experience stable unemployment.

We next consider the transition rates for groups with different levels of educational attainment. These are shown in Fig. 4 below. The decomposition by education level is striking for the visual pattern that it displays. As expected, we find a clear and strong gradient in terms of labour market outcomes by education level. More than two thirds of respondents with less than a matric who are unemployed in wave 1 are also unemployed in waves 2 and 3. This number is about 50% for those with a matric, and decreases to 37.8% amongst those with some tertiary qualification. Note that the 37.8% is only relatively small, it still implies that more than 1 in 3 respondents with some tertiary education are in stable unemployment, which attests to the difficulty in finding jobs in the SA labour market as a whole. Of the unemployed in wave 1, there is also a clear educational gradient in terms of the likelihood of finding stable employment relative to unstable employment.

When we look at employment stability amongst those who were employed in wave 1, we again see the value of higher levels of education. Only 54.5% of those with less than a matric are stably employed, while close to 1 in 5 enter stable unemployment. The corresponding job stability numbers are 70.5% and 78.8% for those with a matric and those with some tertiary education, and the likelihood of entering stable unemployment are less than 1 in 10 and 1 in 25 for people with these qualifications respectively.

Education thus helps people to find employments, and moreover, has a strong effect in terms of the stability of one's employment status that one can expect in the types of jobs that better skilled people do.

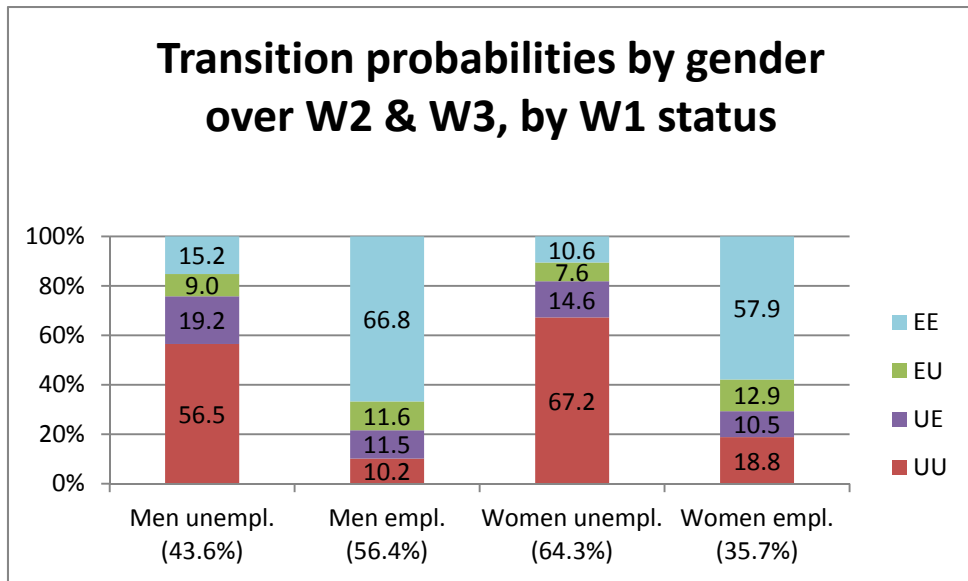**Transition probabilities by gender over W2 & W3, by W1 status**

Transition probabilities by educational categories over W2 & W3: By W1 status

Transition probabilities in Urban & Rural areas over W2 & W3 status: By W1 status

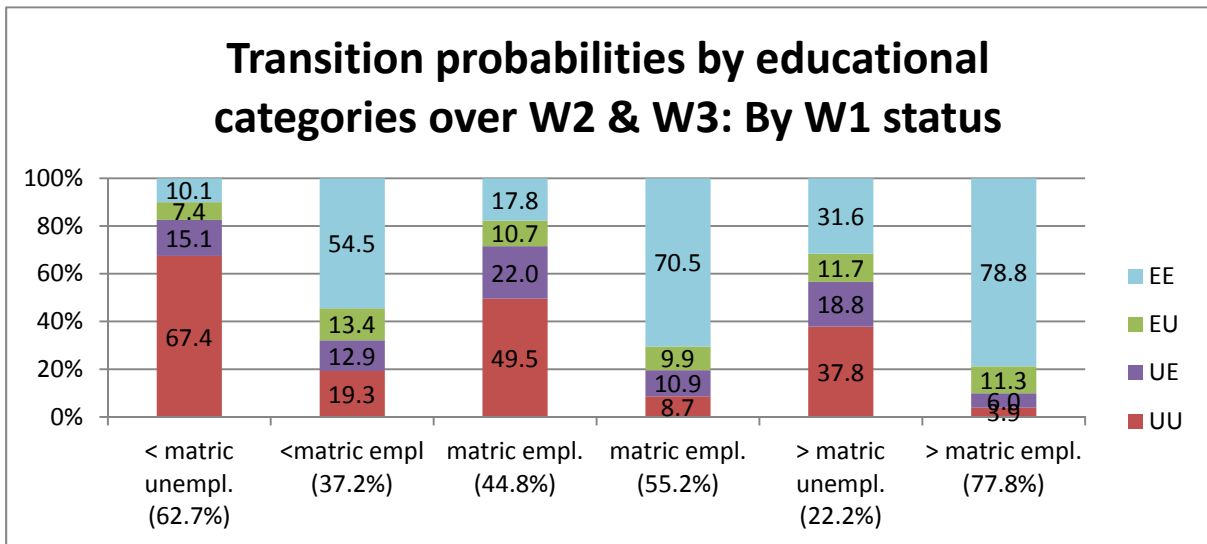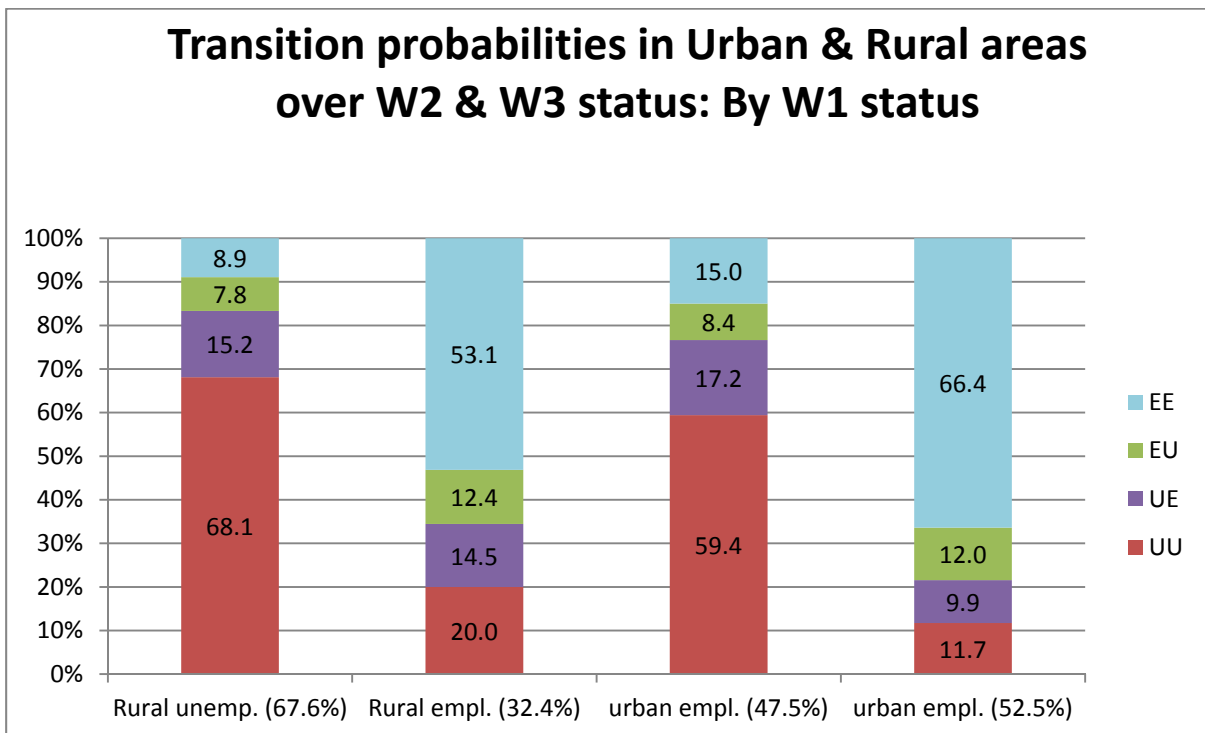| Category	Rural unemp. (67.6%)	Rural empl. (32.4%)	urban empl. (47.5%)	urban empl. (52.5%)
EE	8.9	53.1	15.0	66.4
EU	7.8	12.4	8.4	12.0
UE	15.2	14.5	17.2	9.9
UU	68.1	20.0	59.4	11.7

Our final graph shows the transition rates for people who resided in rural and urban areas in wave 1 separately. As expected, the labour market opportunities in rural areas seem to present fewer opportunities than those in urban areas. Note that migration might reduce the differentials as our classification of 'urban' is based on wave 1 location only. As with the other graphs, it is certainly better to be in an urban environment if you want to find employment or keep employment. Here, we need to be cautious because a disproportionate number of youth and older adults live in rural areas, and in the case of retirees and students it may be that the differential between 'unemployment' and 'not economically active' is a meaningful one.

Measures of earnings volatility

In Table four below, we present the mean standard deviations and coefficients of variation for each of our groups. As discussed, an expected finding is that groups with higher wages will have a higher standard deviation in their earnings on average, but not necessarily a higher average coefficient of variation.

The mean overall standard deviation is large, at R1433. Given that mean earnings in this sample is between R2100 and R2600 per month in the panel, the standard deviation lies between 50% and 66% of the mean, depending on which wave's mean we are comparing it to. This highlights that there is substantial volatility in earnings in the SA labour market.

Amongst the different age groups, the mean standard deviation in earnings increases as age increases. Here it is useful to also compare with the mean coefficient of variation (CV), as

older people earn substantially more than younger ones. Thus, the pattern is reversed once we use the CV measure. Taken together, we can say that older people experience greater earnings volatility, but a large fraction of this is because they have higher wages when they are employed.

Africans experience less volatility than the overall sample when measured by the standard deviation, but similar volatility as the overall sample when using the CV measure. This is consistent with the earlier observation that mean African wages are lower than in the overall sample.

Men have greater levels of volatility using both measures. So while men have better employment prospects than women, and better earnings on average conditional on employment, the aggregate effect is a higher rate of earnings volatility.

When we look at the standard deviation measure by education, we find a strong positive gradient in terms of volatility with respect to education. This is partly driven by the higher wages that better educated people receive. Using the CV measure, which adjusts for the levels of wages that individuals earn, we find an "inverted-U" shape. Thus, the people with a matric have the highest CV value on average, and the ones with less than matric and those with some tertiary qualification have similar but lower levels of volatility. This inverted-U shape reflects partly the employment transitions. People with less than a matric experience lower levels of earnings volatility as they are likely to get trapped in unemployment, which entails no earnings. This reduces the variability in their earnings as a group. People with tertiary levels of education tend to remain in stable levels of employment, and the variability in one's earnings if one is stably employed is much smaller than if one is in unstable employment.

Finally, we observe that earnings volatility is lower in rural areas using both of the measures that we calculated.

Table 4: Earnings volatility: Mean Std. Deviation and coeff. of variation
(Within person)

	Mean std. Dev	Mean CV
Entire Sample	1433.3	0.641
Age distribution		
% youth (16 - 29)	890.0	0.660
% prime aged (30 - 49)	1718.5	0.657
% Older (50 - 64)	2050.6	0.550
Race		
% African	1037.3	0.645
Gender		
% male	2032.9	0.683
% female	915.8	0.603
Education		
% < matric	716.1	0.625
% matric	1983.9	0.703
% some tertiary	4925.0	0.629
Location		
% Rural	687.4	0.603
% Urban	1849.5	0.661

Notes:
1. People with no earnings were assigned a CV value of 0.

Regression results

One of the challenges in the analysis thus far has been that the groups are not independent of each other. For example, better educated people are also more likely to be young and to reside in cities. We next estimate a standard OLS regression model with all of the group identifiers included simultaneously. This allows us to obtain a measure of the correlation between a demographic characteristic and the volatility measure, under the assumption that the values of all the other variables are held constant. The results from our regressions are presented in Table 5 below.

The first thing to note is that all the variables that we include are highly significant in the standard deviation regression. This is unsurprising since we have already established that these groups experience different wage distributions and employment prospects. Older people, men and those with a matric or a tertiary qualification have large and and positive coefficients.

The regression with the CV as the dependent variable shows a slightly different picture. Only the male, matric and urban coefficients are statistically significant and positive. This corroborates the insights obtained from the table of means above. The older adults experience significantly lower levels of earnings volatility, relative to youth. Part of this reflect job stability for some older adults, and part of this reflects the movement into retirement, which will entail zero volatility if a retiree was not employed in each of the three waves.

Table 5: Regressions on Standard Deviation and Coefficient of Variation

	Std. Dev.		Coef. of Var.	
	Coef.	t - stat	Coef.	t - stat
prime	508**	5.27	-0.0008	-0.03
older	1008.9**	3.52	-0.0957**	-3.13
african	-1124**	-4.71	0.0273	0.91
male	1146.9**	8.12	0.0765**	3.71
matric	1228.5**	6.74	0.0591**	2.12
tertiary	3749.8**	8.68	-0.0034	-0.1
urban_w1	438.0**	7.17	0.0578**	2.79
_cons	484.3**	2.34	0.5512**	14.58
N	9016		9016	
R-squared	0.1798		0.01	

Notes
1. Robust standard errors were calculated.
2. People with no earnings in any of the three waves were assigned a CV value of 0.
3. An * denotes significance at the 5% level, ** at the 1% level

5. Conclusion

In this paper, we set out to obtain a measure of the degree of earnings volatility in South Africa. We made use of the first three waves of the nationally representative National Income Dynamics Study data. The first set of analyses we undertook was to estimate the flows into and out of employment. Consistent with previous literature, we observed that there is a relatively high level of churning in the SA labour market. Given this, it is not surprising that there is also a substantial amount of volatility in the earnings of respondents who survive into the balanced panel. The mean standard deviation in earnings across the three waves lies between 50% and 66% of the mean earnings depending on the time period, and the mean coefficient of variation in earnings is 0.641.

When we estimated the differences in volatility across different groups, where the groups were defined by age, race, gender, educational attainment and geography, we observed substantial differences along these dimensions. In our regression analyses, the results differed depending on whether one used the standard deviation or the coefficient of

variation as the dependent variable. Using the coefficient of variation, which adjusts for the mean level of earnings within respondents, we found that older adults experience less earnings volatility, while men and respondents in urban areas experienced more volatility. People with only a matric level of education experienced more volatility than both people who had not finished high school as well as people who had some amount of tertiary education.

References

Cichello, P. L., Fields, G. S., & Leibbrandt, M. (2005). Earnings and employment dynamics for Africans in post-apartheid South Africa: A panel study of KwaZulu-Natal. *Journal of African Economies*, *14*(2), 143-190.

Gottschalk, P., Moffitt, R. (1994). The growth of earnings instability in the US labor market. *Brookings Papers on Economic Activity*, (2), 217-254.

Hacker, J. S. (2006). The Great Risk Shift: The Assault on American Jobs, Families. *Health Care, and Retirement and How You Can Fight Back. New York: Oxford University Press*.
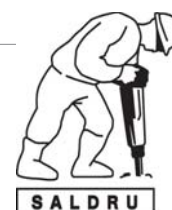
Leibbrandt, M., Bhorat, H. & Woolard, I. (2001), Understanding Contemporary Household Inequality in South Africa, Chapter 2 in Bhorat, H., Leibbrandt, M., Maziya, M., van der Berg, S. & Woolard, I. (2001). Fighting poverty: Labour markets and inequality in South Africa. UCT Press: Cape Town.

Shin, D., & Solon, G. (2011). Trends in men's earnings volatility: What does the Panel Study of Income Dynamics show?. *Journal of Public Economics*, *95*(7), 973-982.

southern africa labour and development research unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.