



Estimating income mobility when income is measured with error: The case of South Africa

Rulof P. Burger, Stephan Klasen and Asmus Zoch

ERSA working paper 607

May 2016

Economic Research Southern Africa (ERSA) is a research programme funded by the National Treasury of South Africa.

The views expressed are those of the author(s) and do not necessarily represent those of the funder, ERSA or the author's affiliated institution(s). ERSA shall not be liable to any person for inaccurate information or opinions contained herein.

Estimating Income Mobility When Income is Measured with Error: The Case of South Africa*

Rulof P. Burger[‡] Stephan Klasen[§] and Asmus Zoch[¶]

May 19, 2016

Abstract

There are long-standing concerns that household income mobility is over-estimated due to measurement errors in reported incomes, especially in developing countries where collecting reliable survey data is often difficult. We propose a new approach that exploits the existence of three waves of panel data to can be used to simultaneously estimate the extent of income mobility and the reliability of the income measure. This estimator is more efficient than 2SLS estimators used in other studies and produces over-identifying restrictions that can be used to test the validity of our identifying assumptions. We also introduce a nonparametric generalisation in which both the speed of income convergence and the reliability of the income measure varies with the initial income level. This approach is applied to a three-wave South African panel dataset. The results suggest that the conventional method over-estimates the extent of income mobility by a factor of more than 4 and that about 20% of variation in reported household income is due to measurement error. This result is robust to the choice of income mobility measure. Nonparametric estimates show that there is relatively high (upward) income mobility for poor households, but very little (downward) income mobility for rich households, and that income is more reliably captured for rich than for poor households.

KEYWORDS: Income mobility, inequality, longitudinal data analysis, measurement error

JEL CODES: J62, D63, C23

*The authors would like to thank participants at the Centre for the Study of African Economies conference, the Micro-econometric Analysis of South African Data conference, the DataFirst Data Quality Workshop, and departmental seminars at Stellenbosch University and the University of Cape Town, as well as useful comments from Martin Abel. All remaining errors remain the responsibility of the authors.

[†]Economics Department, University of Stellenbosch, South Africa.

[‡]Centre for Studies of African Economics, University of Oxford, United Kingdom.

[§]Economics Department, University of Göttingen, Germany.

[¶]Centre for Studies of African Economics, University of Oxford, United Kingdom.

1 Introduction

With the increasing availability of panel data in developing countries, studying economic mobility is now feasible and has been done in an increasing number of studies. This is often done in the setting of a so-called “micro growth regression”, where income growth is regressed on initial income and some other covariates (e.g. Fields, Cichello, Freije, Menéndez, & Newhouse, 2003a; Woolard & Klasen, 2005). A robust finding of that literature is a rather large negative and highly significant coefficient on the initial income variable, indicating high mobility and suggesting a high speed of “beta-convergence”. For example, Fields et al. (2003a) estimate a convergence coefficient of -0.56 (over 5 years) for South Africa, which suggests that one should expect half the income gap between the richest and poorest household to be eliminated every 4.3 years. However, there are also long-standing concerns that micro mobility is over-estimated due to errors in income measures, especially in developing countries where collecting reliable survey data is often difficult. Indeed, Fields (2008b) acknowledges that “a task for the future is to estimate empirically the effect of measurement error on estimates of . . . micro-mobility”. Existing approaches to address this issue tend to use instrumental variable (IV) approaches to instrument initial incomes, but the suitability of the instruments as well as the reliability and robustness to these approaches is open to question (e.g. Woolard and Klasen (2005), Fields, Cichello, Freije, Menéndez, and Newhouse (2003b)).

In this paper we develop an alternative approach that exploits the existence of three waves of panel data to simultaneously estimate the speed of convergence and the extent of measurement error. Our estimates are more efficient than the two-stage least squares (2SLS) estimator that have been used in other studies and can be generalised to allow both the speed of income convergence and the reliability of the income measure to vary with the initial level of income. This approach is applied to a three-wave South African household panel dataset. Studying income mobility in South Africa is particularly pertinent as the economy and society has undergone significant changes since the end of apartheid in 1994. Increasing incomes of previously disadvantaged poor population groups was a key policy target of successful governments led by the African National Congress (Van der Berg, Burger, Burger, Louw, & Yu, 2005; Woolard & Klasen, 2005; Özler, 2007). Our results suggest that previous studies have over-estimated the extent of income mobility by a factor of between 4 and 5 and that about 20% of variation in reported household income is due to measurement error. We demonstrate that the same effect is observed for an alternative measure of income mobility: Shorrocks’ rigidity index. Nonparametric estimates show that there is nevertheless relatively high (upward) income mobility for poor households, but very little (downward) income mobility for rich households, and that income is more reliably captured for rich than for poor households.

2 Income mobility and measurement error

There are various ways to measure the mobility of households in the income distribution¹. Borrowing from the macroeconomic convergence literature (e.g. Barro and Sala-i-Martin (1992), Quah (1996)), we will restrict our attention in this paper to the concept of weak unconditional beta convergence. Accordingly, the log per capita household income, y_t^* , is characterised as an autoregressive process of order 1, or AR(1) process:

$$y_t^* = \mu + \rho y_{t-1}^* + u_t$$

Current income depends on past income as well as a stochastic income shock, u_t which is often assumed to be $iid(0, \sigma_u^2)$ (Fields, 2008a, p. 5). The proportional change in income between two periods can then be expressed as

$$\Delta y_t^* = y_t^* - y_{t-1}^* = \mu + \beta y_{t-1}^* + u_t \quad (1)$$

where $\beta \equiv \rho - 1$ reflects the extent of income mobility in the economy. This specification is deliberately parsimonious², since the object of interest is the speed of income convergence rather than the causal mechanism that determines household income. Most studies (e.g. Jarvis and Jenkins (1998), Fields et al. (2003a), Antman and McKenzie (2007b), Fields, Duval-Hernández, Freije, and Puerta (2014)) have focussed on testing $\beta = 0$ against the alternative hypothesis $\beta < 0$. If $\beta = 0$ then there is no tendency for rich and poor household to experience different growth rates, whereas if $\beta < 0$, then poor households tend to grow more rapidly than rich ones. The empirical literature on unconditional convergence in developing countries produces a “virtual consensus” (Fields, 2008a, p. 6) that poorer households’ incomes grow more quickly than those of richer ones. Evidence of weak unconditional income convergence has been established, among others, for Indonesia, Venezuela, South Africa (Fields et al., 2003a; Woolard & Klasen, 2005), Vietnam (Glewwe, 2012), Argentina, Mexico (Fields et al., 2014) and China (Heng, Shi, & Quheng, 2006; Khor & Pencavel, 2006).

Most empirical studies focus mainly on whether there is evidence of income convergence, but the point estimate of β can also be used to gauge the speed at which this convergence occurs. In the cross-country growth literature it is sometimes insightful to calculate how rapidly countries converge on their steady states (e.g. Barro and Sala-i-Martin (2004, p. 58)). A related approach allows us to calculate how rapidly the income gap between two randomly chosen households will tend to disappear. Equation [1] and the assumption that income shocks are i.i.d. imply that the expected one-period change in the relative income gap between any two households (denoted A and B) can be expressed

¹ See Jäntti and Jenkins (2015) for a recent overview of this literature.

² A different strand of the literature focuses on conditional income convergence: the speed at which households converge on their own expected income levels, as determined by their observable covariates or household fixed effects.

as

$$\frac{(y_{A,t-1}^* - y_{B,t-1}^*) - E(y_{A,t}^* - y_{B,t}^* | y_{A,t-1}^* - y_{B,t-1}^*)}{y_{A,t-1}^* - y_{B,t-1}^*} = -\beta$$

In other words, if $\beta < 0$ then $-\beta$ represents the share of any income gap that we would expect to be eliminated between periods $t - 1$ and t . This convergence parameter can be used to calculate the expected half-life of an income gap (i.e. the expected duration required for half of any income gap to be eliminated) as $t \doteq \frac{0.69}{\log(1+\beta)}$ periods. For example, Fields et al. (2003a) find convergence coefficients of -0.56 (over 5 years for South Africa), -0.53 (over 4 years for Indonesia), -0.52 (over 1 year for Spain) and -0.64 (over 1 year for Venezuela). These coefficients imply that the expected half-life of the income gap between the richest and poorest households is 4.3 years (South Africa), 3.7 years (Indonesia), 1 year (Spain) and 0.7 years (Venezuela), respectively.

Of course, the estimate of β is only informative about the extent of income mobility if such an estimate is reliable. In practice, income measures obtained from surveys are usually only noisy approximations of true household income, especially in developing countries where collecting reliable survey data can be difficult. In fact, many of the above-mentioned empirical studies mention the issue of measurement error as a potential confounding factor that may lead to an over-estimation of the extent of income mobility and many try to address this issue econometrically.

In order to formally investigate the effect of measurement error, suppose the available income measure, y_t , suffers from classical measurement error, so that $e_t \equiv y_t - y_t^* \sim iid(0, \sigma_e^2)$. Rewriting equation [1] in terms of the observed but noisy income measure gives:

$$\Delta y_t = \mu + \beta y_{t-1} + u_t + e_t - (\beta + 1)e_{t-1} \quad (2)$$

The econometric problem is that initial income y_{t-1} is negatively correlated with the model error term via the initial period measurement error term e_{t-1} , which will downwardly bias the OLS estimate of the convergence parameter β . Under the maintained assumptions that both e_t and u_t are i.i.d., the expected value of the OLS slope coefficient obtained from regressing Δy_t on y_{t-1} (which we denote as θ_1) can be expressed as:

$$E(\theta_1) = \frac{Cov(\Delta y_t, y_{t-1})}{Var(y_{t-1})} = \beta \frac{-(\beta + 1)\sigma_e^2}{Var(y_{t-1})} = (\beta + 1)\alpha - 1 \quad (3)$$

where $\alpha \equiv \frac{Var(y_{t-1}^*)}{Var(y_{t-1})} = \frac{Var(y_{t-1}^*)}{Var(y_{t-1}^*) + \sigma_e^2}$ is the share of the total variation in the initial income measure that is due to variation in actual initial income, y_{t-1}^* , rather than measurement error, e_{t-1} . This parameter is sometimes referred to as the “reliability statistic” (Gottschalk & Huynh, 2010; Abowd & Stinson, 2013). It is restricted to lie within the unit interval and represents the reliability of the observed measure of initial income y_{t-1} . A value of $\alpha = 1$ represents the case of income measured without error, whereas $\alpha = 0$ would indicate that the income measure is all noise and contains no information about actual household

income. In the case of no measurement error it follows from equation [3] that $E(\theta_1|\alpha = 1) = \beta$, so the OLS estimator will provide an unbiased estimate of the extent of income mobility. However, whenever income is measured with some error equation [3] indicates that $E(\theta_1|\alpha < 1) < \beta$. This will create the appearance of income mobility, even where none exist. Intuitively, if household income is reported with error (and this error is uncorrelated over time), then we would expect households who under-reported their income in the previous period to report a higher income in the current period, and vice versa, even if their actual household income was unchanged.

The most common way of addressing measurement error in income mobility studies is to use instrumental variables to obtain a predicted value of lagged income in equation [2] (Fields et al., 2003b; Newhouse, 2005; Lee, 2009; Glewwe, 2012; Fields et al., 2014). Glewwe (2012) finds that that at least 15%, and perhaps as much as 42%, of estimated mobility in Vietnam is due to measurement error bias. Turning to previous studies on South African income dynamics, Agüero, Carter, and May (2007) instrument for initial income using household health measures and find that measurement error accounts for between 14% and 60% of all mobility between two successive waves of the South African Kwazulu-Natal Income Dynamics Study (KIDS) panel dataset. Woolard and Klasen (2005) apply a similar approach to the same dataset but find that their results are largely unaffected by measurement error. Lechtenfeld and Zoch (2014) use a three wave panel dataset to instrument for initial income with previous period income and conclude that conditional income convergence is over-estimated by 39% in the KIDS panel and by 77% in the South African National Income Dynamics Study (NIDS) panel dataset.

A number of studies have used validation data from different sources, like administrative records, to investigate directly the reliability of self-reported labour market earnings data (Bound & Krueger, 1991; Bound, Brown, & Mathiowetz, 2001; Gottschalk & Huynh, 2010; Akee, 2011). Abowd and Stinson (2013) find that self-reported earnings of US workers have a reliability statistic of 0.7. When they omit all imputed earnings from the sample, this ratio increases to 0.78. In the only developing country validation study that we are aware of Akee (2011) estimates reliability statistics of between 0.42 and 0.7 for self-reported earnings of workers in Micronesia. These estimates could serve as useful benchmarks for our own estimate of the reliability statistic for per capita household income.

Validation studies typically find that the measurement error in earnings is serially correlated over time and negatively correlated to true earnings. In this more general case, the expected value of θ_1 can be more accurately expressed as:

$$E(\theta_1) = (\beta + 1)\alpha - 1 + \frac{Cov(e_t, y_{t-1}^*) + Cov(e_t, e_{t-1}) - (\beta + 1)Cov(e_{t-1}, y_{t-1}^*)}{Var(y_{t-1})} \quad (4)$$

Antman and McKenzie (2007b) argue, based on the insights from validation studies, that the last term on the RHS of equation [4] will be positive and so the tendency of classical measurement error to overstate earnings mobility could be

partly been offset by the non-classical features of this error.

The finding that the measurement error in self-reported earnings reveals some non-classical features should make us wary of uncritically applying the assumption of classical measurement error to an analysis of per capita household income dynamics. However, validation data is rarely available for developing country surveys or for per capita household income, so we are constrained in what we can do with existing data. Furthermore, if the household size variable that is used to scale total household income also suffers from mean-reverting measurement error – as is typical for categorical variables (Bound et al., 2001, p. 3725) – then this will tend to reduce the correlation between per capita income and its measurement error. This means that the classical measurement error assumption could still offer a useful first approximation that can be used to scrutinise the effects of measurement error in the absence of better data or techniques that require less restrictive identifying assumptions.

One recently popular approach that attempts to address non-classical measurement error with the available data makes use of pseudo panels (Antman & McKenzie, 2007a; Antman & McKenzie, 2007b; Cuesta, Ñopo, & Pizzolitto, 2011). Successive cross-sectional datasets are used to track the average income for households with heads from the same birth cohort over time. The benefit of this approach is that the within-cohort averaging procedure will remove the effects of income measurement error in sufficiently large cohorts, even where this error is non-classical in nature. Unfortunately, it also averages away all of the highly informative within-cohort variation in household income, which will dramatically reduce the estimator precision and make the estimates highly vulnerable to any deviations from its identifying assumptions. Fields and Viollaz (2013) apply pseudo-panel estimators to actual panel data, and find that these methods perform poorly in predicting the actual income mobility pattern.

Although validation data is rarely available in developing countries, it is increasingly common to have three consecutive waves of panel data with which to study household income dynamics. In such cases there is additional information that can be used to distinguish between true income mobility and measurement error. The remainder of this paper will develop an approach to do exactly that. We start by assuming that self-reported per capita household income suffers from classical measurement error, but in section 4.4 we relax this assumption by allowing the reliability statistic to vary across income levels.

3 Regression coefficients in a three wave panel data set

As soon as we have more than two waves of panel data, we are required to make additional assumptions about how income mobility and measurement error changes between waves. The income dynamics equation [1] can be generalised as

$$\Delta y_t^* = \mu_t + \beta_t y_{t-1}^* + u_t \tag{5}$$

in which both the intercept and the slope of the first-order autoregressive income process are time-varying. Our proposed approach requires assuming that $\beta_t = \beta < 0$ and u_t . The income convergence coefficient is therefore assumed to be constant over the period under consideration. Given the relatively short periods studied in most of this literature, the market forces and institutional determinants of income mobility are unlikely to have changed substantially. As we show below, it is possible to empirically test the validity of this assumption.

The income intercept term is completely unrestricted over time, which allows income to follow a potentially non-linear time trend represented by the u_t parameters. We also maintain the assumption that income measurement error is classical: $e_t \equiv y_{t-1}^* \sim \text{nid}(0, \sigma_e^2)$. It is possible to formally test whether our parameter estimates are consistent with this version of the model and in section 4.4 we also consider one particular form of non-classical measurement error. It is possible to further relax some of these assumptions by, for example, allowing β , σ_u^2 or σ_e^2 to change between waves, but this comes at the cost of losing over-identifying restrictions and estimator precision. Where more than three waves of panel data are available there is even more scope for imposing less restrictive identifying assumptions. However, in this paper we will restrict our attention to the more restrictive specification in a three wave panel dataset.

In this case there are at least seven regression coefficients that can be used to inform our estimates of the convergence and income measure reliability parameters, β and α . These coefficients are all easy to estimate and straightforward to interpret. Let L denote the linear projection operator so that, for example, $L(y_2|y_1)$ represents the linear projection of y_2 on y_1 . The seven regression coefficients are defined in the first column of Table 1³ and discussed below.

The first coefficient, θ_1 , represents the effect of wave 1 income, y_1 , on subsequent income growth between waves 1 and 2, Δy_2 . We define θ_2 as the same relationship between wave 2 income, y_3 , and income growth between waves 2 and 3, Δy_3 . These two coefficients represent the conventionally reported estimates of the convergence parameter in a two-wave panel dataset, and either coefficient should provide a consistent estimate of β if income is measured without error⁴.

As discussed in section 2 and elsewhere in this literature, measurement error will tend to bias the estimates of θ_1 and θ_2 away from β and towards -1, since $E(\hat{\theta}_1|\beta, \alpha) = E(\hat{\theta}_2|\beta, \alpha) = (\beta + 1)\alpha - 1$. One way to gauge the reliability of θ_1 or θ_2 as estimates of β is to compare them to regression coefficients $\hat{\theta}_3$ (the regression coefficient obtained from regressing Δy_2 on y_1) and $\hat{\theta}_4$ (obtained from regressing $y_3 - y_1$ on y_1). In the absence of measurement error, a stationary AR(1) process that eliminates in expectation $-\beta$ of income gaps between waves 1 and 2 should eliminate a smaller proportion $-\beta(\beta+1)$ of the initial income gaps between waves 2 and 3. Between waves 1 and 3 the total proportional income convergence should therefore be $-\beta(\beta + 2)$. In the absence of measurement

³ Assuming that y_t is demeaned allows us to omit the intercept term and is without loss of generality.

⁴ Regardless of whether or not income is measured with error, a comparison of the estimates of regression coefficients θ_1 and θ_2

allows a test of the assumption that $\beta_t = \beta$.

error, regression coefficients θ_3 and θ_4 provide estimates of these two quantities: $E(\hat{\theta}_3|\beta, \alpha = 1) = \beta(\beta + 1)$ and $E(\hat{\theta}_4|\beta, \alpha = 1) = \beta(\beta + 2)$.

However, if the data is measured with classical error, then $E(\hat{\theta}_3|\beta, \alpha) = \alpha\beta(\beta + 1)$ and $E(\hat{\theta}_4|\beta, \alpha) = \alpha\beta(\beta + 1)^2 - 1$. Whereas classical measurement error downwardly biases $\hat{\theta}_1$ and $\hat{\theta}_2$ it will upwardly bias $\hat{\theta}_3$. This is because measurement error of the regressor used to produce $\hat{\theta}_3, y_1$, is uncorrelated to the measurement error of the regressand, and hence it only suffers the usual (upward) attenuation bias. Classical measurement error therefore leads to an over-estimation of income convergence between waves 1 and 2, and an under-estimation of convergence between waves 2 and 3. The effect on total income convergence between waves 1 and 3 is dominated by the former effect, so coefficient $\hat{\theta}_4$ tends to over-estimate income mobility, but less so than coefficient $\hat{\theta}_1$. This offers a natural way of using the coefficient estimates⁵ of θ_3 or θ_4 to test for the presence of measurement error: check whether there is surprisingly little additional income convergence between waves 2 and 3, given the income mobility that is supposedly observed between waves 1 and 2.

Additional information about the convergence parameter is contained in regression coefficients $\hat{\theta}_5$ and $\hat{\theta}_6$, the coefficients on y_1 and y_2 when simultaneously included in a regression of Δy_3 . If income is measured without error then y_1 should have no effect on Δy_3 after we control for y_2 , and the effect of y_2 is simply the convergence parameter β : $E(\hat{\theta}_5|\beta, \alpha = 1) = 0$ and $E(\hat{\theta}_6|\beta, \alpha = 1) = \beta$.

As with the other regression coefficients, the expected values of $\hat{\theta}_5$ and $\hat{\theta}_6$ are affected by measurement error, and in a way that provides us with useful information about income mobility and the reliability of the income measure. Let us start by considering the estimated effect of y_2 on Δy_3 . If we do not control for y_1 then this estimate is represented by $\hat{\theta}_2$, which is known to be a downwardly biased estimate of actual income convergence if income is measured with error. Now, y_1 will be correlated with the true value of wave 2 income⁶, y_2^* , but not with its measurement error, e_2 . This means that controlling for y_1 will exacerbate the bias in the coefficient on y_2 . The estimate of θ_6 will therefore be more downwardly biased than the estimate of θ_2 : $E(\hat{\theta}_6|\beta, \alpha) = \frac{1 - \alpha(\beta + 1) + \alpha^2\beta(\beta + 1)^2}{\alpha^2(\beta + 1)^2 - 1}$. Because of this bias the correlation between y_2 and Δy_3 is not fully accounted for in the relevant regression coefficient, which means y_1 should reveal an opportunistic positive correlation with Δy_3 . More specifically, measurement error should upwardly bias regression coefficient θ_5 so that $E(\hat{\theta}_5|\beta, \alpha) = \frac{(\beta + 1)^2(\alpha - 1)\alpha}{\alpha^2(\beta + 1)^2 - 1}$. Measurement error will therefore make an AR(1) process seem like an AR(2) process in which income growth depends negatively on the first lag of income and positively on second lag of income. This provides another test of the validity of our model⁷.

⁵Note that regardless of the values of β and α , $\theta_1 + \theta_3 = \theta_4$. This means that these two regressions coefficients only add one additional linearly independent population moment that can be used to test hypotheses about or estimate the values of β and α .

⁶This is true unless $\beta = -1$.

⁷Coefficients θ_1, θ_3 and θ_5 are linearly dependent, so θ_5 on its own does not add any new information to the model. θ_6 , on the other hand, is linearly independent of the other regression coefficients, and can therefore provide new information to test or estimate the model

Finally, θ_7 is defined as the slope coefficient obtained from regressing Δy_3 on Δy_2 . In the absence of measurement error this coefficient estimate has expected value $E(\hat{\theta}_7|\beta, \alpha = 1) = \frac{1}{2}\beta$, which captures the fact that households that experienced more rapid income growth between waves 1 and 2 should expect to experience slower subsequent income growth. If we allow for measurement error then the expected value of this coefficient estimate becomes $E(\hat{\theta}_7|\beta, \alpha) = -\frac{1-\alpha+\alpha\beta^2}{2(1-\alpha-\alpha\beta)}$. If income is measured with error, then the negative correlation between Δy_2 and Δy_3 should be larger than expected in the no measurement error case.

4 Estimation and hypothesis testing

4.1 Informal approaches

There are at least two hypotheses that we may be interested in testing. First, is income measured without error: $\alpha = 1$? Secondly, do our maintained assumptions of classical measurement error and a first-order autoregressive income process produce an internally consistent set of regression coefficients? There are various ways to use the above-defined regression coefficients to explore the validity of these two hypotheses.

We start by considering the first hypothesis. Under the maintained assumptions consistent estimates of (α, β) can be easily obtained by combining any one of a number of pairs of coefficients. For example⁸, estimates of θ_1 and θ_3 can be used to estimate the parameters of interest as:

$$\hat{\beta} = \frac{\theta_3}{\hat{\theta}_1 + 1} \text{ and } \alpha = \frac{(\hat{\theta} + 1)^2}{\hat{\theta}_3 + \hat{\theta}_1 + 1} \quad (6)$$

The resulting estimate of α can be used to test directly the hypothesis of no measurement error⁹. Of course, the remaining regression coefficients provide additional information that can be used to assess the validity of this hypothesis. A straightforward but informal test¹⁰ that uses all of the regression coefficients would be to combine the assumption that $\alpha = 1$ with the sample estimate $\hat{\theta}_1$ (as an estimate for $\hat{\beta}$) to calculate the predicted values of the other six regression coefficients using the equations for $E(\hat{\theta}_K|\beta = \hat{\theta}_1, \alpha = 1)$ in column 3 of Table 1. If income is measured without error and the maintained assumptions of a first-order autoregressive income process is valid, then the predicted and estimated coefficient values should only differ due to sampling variation. However, if the maintained assumptions are violated then this may cause the estimated regression coefficients to be very different from the predicted values. Of course, such differences can also arise if the other maintained assumptions are false.

parameters.

⁸This produces a point estimate β that is identical to the 2SLS estimate of the effect of y_2 on Δy_3 in which y_1 is used to instrument y_2 .

⁹The standard error of this estimate can be approximated with the delta method.

¹⁰A more formal test of this hypothesis is discussed in section 4.2 below.

The second hypothesis – that these other maintained assumptions are valid – can be tested indirectly by comparing the estimated regression coefficients to the predicted values obtained using estimates of β and α and the more general equations for $E(\hat{\theta}_K|\beta, \alpha)$ in column 4 of Table 1. In section 5 below, this approach is applied to South African household per capita income data. We find that the regression coefficients are indeed very similar to the values predicted under the assumption of classical measurement error, and very different from the values predicted under the assumption of no measurement error.

A different but related approach would be to use the seven estimated regression coefficients to derive implied estimates of β . If income is measured without error then the expected values of $E(\hat{\theta}_K|\beta, \alpha = 1)$ in column 3 of Table 1 can be used to calculate the values of β implied by each¹¹ of the regression coefficients. If these estimates are observed to lie within a relatively narrow range, then this provides evidence in support of the hypothesis of no measurement error. More generally, an estimate of α could be used with the expected values of the regression coefficients for $E(\hat{\theta}_K|\beta, \alpha = \hat{\alpha})$ in column 4 of Table 1 to calculate the values of β implied by each of the regression coefficients if the data is measured with error. If these estimates all lie within a narrow range, whereas those obtained under the no measurement error assumption do not, then it provides evidence against the assumption of measurement error, but no evidence against the maintained assumptions of classical measurement error and a first-order autoregressive income process. These two sets of estimates for the South African data are reported in rows 4 and 5 of Table 3. The estimates demonstrate that the observed regression coefficients are unlikely to have been produced by the same value of β if income is measured without error, whereas allowing for classical measurement error is sufficient to produce implied values of β that lie within a very narrow range.

4.2 GMM approach

A system estimator offers a more efficient approach to estimating the model parameters and testing the over-identifying restrictions than the informal approach outlined above. In the presence of classical measurement error there are five linearly independent coefficients that depend on two unknown parameters. The generalised method of moments (GMM) estimator provides a (possibly asymptotically efficient) way of estimating the values of β and α . Given the relationship between our parameters of interest (β, α) and the vector of regression coefficients $\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7]$, we can construct a vector of sample

¹¹With the exception of θ_5 which has an expected value of 0 if income is measured without error.

moments:

$$g(y_{it}, \theta(\beta, \alpha)) = \begin{bmatrix} (y_{2i} - y_{1i} - \theta_1 y_{1i}) y_{1i} \\ (y_{3i} - y_{2i} - \theta_2 y_{2i}) y_{2i} \\ (y_{3i} - y_{2i} - \theta_3 y_{1i}) y_{1i} \\ (y_{3i} - y_{2i} - \theta_5 y_{1i} - \theta_6 y_{2i}) y_{2i} \\ (y_{3i} - y_{2i} - \theta_7 (y_{2i} - y_{1i})) (y_{2i} - y_{1i}) \end{bmatrix}$$

The identifying assumption $E[g(y_{it}, \theta(\beta_0, \alpha_0))] = 0$ follows directly from the assumptions that both u_t and e_t are i.i.d. processes. The GMM estimator can then be expressed as:

$$(\tilde{\beta}, \hat{\alpha}) = \arg \min_{\beta, \alpha} \left(\frac{1}{N} \sum_{it} g(y_{it}, \theta(\beta, \alpha)) \right)' \hat{w} \left(\frac{1}{N} \sum_{it} g(y_{it}, \theta(\beta, \alpha)) \right)$$

where \hat{w} is the weighting matrix. Assuming that this is an optimal weighting matrix and that our over-identifying restrictions are valid implies that

$$I = \left(\frac{1}{\sqrt{N}} \sum_{it} g(y_{it}, \theta(\beta, \alpha)) \right)' \hat{w} \left(\frac{1}{\sqrt{N}} \sum_{it} g(y_{it}, \theta(\beta, \alpha)) \right)$$

has a chi-squared limiting distribution with three degrees of freedom. This provides a straightforward test of the validity of our identifying assumptions. We can also estimate the value of β under the assumption that $\alpha = 1$ to test whether the hypothesis of no measurement error is consistent with all the observed sample regression coefficients. Even in the case of no measurement error, this approach should provide more efficient estimates of the convergence parameter β than estimates of θ_1 and θ_2 .

4.3 Shorrock's index

Of course, the autoregressive income convergence parameter is only one measure of income mobility. An alternative is Shorrock's rigidity index, which measures income rigidity by estimating how much the income distribution is equalised as the period under consideration increases. For example, the T period rigidity index (for income measure y) would be

$$R_T = \frac{I\left(\frac{1}{T} \sum_{t=1}^T y_t\right)}{\sum_{t=1}^T \frac{\bar{y}_t}{\sum_{t=1}^T \bar{y}_t} I(y_t)}$$

where $I(\cdot)$ is some measure of income inequality. The associated mobility index is then simply

$M_T = 1 - R_T$. Our assumptions also allow us to correct estimates of Shorrock's rigidity index for the presence of measurement error if we choose the standard deviation of log per capita household income as our measure of inequality. In the case of no measurement error and our previously maintained identifying assumptions, the expected values of these mobility indices can be related to our

convergence parameter β as $E(\hat{M}_2|\beta, \alpha = 1) = 1 - \sqrt{\frac{\beta+2}{2}}$ and $E(\hat{M}_3|\beta, \alpha = 1) = 1 - \sqrt{\frac{3+4(\beta+1)+2(\beta+1)^2}{9}}$. We can also calculate the expected values of these indices in the presence of classical measurement error as

$$E(\hat{M}_2|\beta, \alpha < 1, \sigma_u^2, \sigma_e^2) = 1 - \sqrt{\frac{1}{2}\left(1 + \frac{\sigma_u^2(\beta+1)}{\sigma_u^2 + \sigma_e^2(1 - (\beta+1)^2)}\right)}$$

and

$$E(\hat{M}_3|\beta, \alpha < 1, \sigma_u^2, \sigma_e^2) = 1 - \sqrt{\frac{1}{3}\left(1 + \frac{1}{3} \frac{\sigma_u^2[4(\beta+1) + 2(\beta+1)^2]}{\sigma_u^2 + \sigma_e^2(1 - (\beta+1)^2)}\right)}$$

where σ_u^2 and σ_e^2 denote the error variances of the income shocks and measurement error terms. These variances can be directly estimated using GMM by relating the variances and covariances of (y_1, y_2, y_3) that are used to construct the seven regression coefficients introduced in section 3 as functions of $(\beta, \sigma_u^2, \sigma_e^2)$.

4.4 Nonparametric extension

As discussed in section 2, we may be concerned that the assumption of classical measurement error is overly restrictive as a basis for identifying the income convergence coefficient. There are many ways in which income measurement error can depart from the assumptions of classical measurement error. One such a deviation occurs when the reliability of the income measure varies with the level of initial income: $\alpha(y_{t-1}^*)$. In this case respondents still provide noisy but unbiased estimates of their household income, but the variance of the measurement error may be larger or smaller for households with higher incomes. Once we relax the assumption that α is constant, it is straightforward to also allow the convergence parameter to vary by initial income: $\beta(y_{t-1}^*)$. In this case we can rewrite equation [2] as

$$\Delta y_t = \mu + \beta(y_{t-1}^*)y_{t-1} + \mu_t + \sigma(y_t^*)e_t - (\beta(y_{t-1}^*) + 1)\sigma(y_{t-1}^*)e_{t-1} \quad (7)$$

where $e_t - 1$ and e_t have now been standardised to have a standard deviation of 1 and $\sigma(y_{t-1}^*)$ reflects the effect of initial income on the standard deviation of the income measurement error. It follows¹² that

$$\frac{\partial E(\Delta y_2|y_1)}{\partial y_1} \cong (\beta(y_1) + 1)\alpha(y_1) - 1 \equiv \theta_1(y_1)$$

$$\frac{\partial E(\Delta y_2|y_1)}{\partial y_1} \cong \alpha(y_1)\beta(y_1)(\beta(y_1) + 1) - 1 \equiv \theta_3(y_1)$$

¹²This follows from the maintained assumptions that $E(u_3|y_1) = E(u_2|y_1) = E(e_3|y_1) = E(e_2|y_1) = 0$, the implication that $E(e_1\sigma_e(y_1^*)y_1) = (1 - E(\alpha(y_1^*)|y_1))y_1$ and the approximations that $E(\alpha(y_1^*)y_1) \cong \alpha(y_1)$, $E(\beta(y_1^*)y_1) \cong \beta(y_1)$ and $\beta(y_1) \cong \beta(y_2)$.

Estimates of these slope parameters can be obtained from local polynomial regressions, and used to estimate the model parameters $(\alpha(y_1), \beta(y_1))$ using a generalisation of equation [6]:

$$\frac{\hat{\theta}_3(y_1)}{\hat{\theta}_1(y_1) + 1} - 1, \frac{(\hat{\theta}_3(y_1) + 1)^2}{\hat{\theta}_3(y_1) + \hat{\theta}_1(y_1) + 1} \quad (8)$$

These estimates, each expressed as a function of y_1 , will provide information about how the reliability of the self-reported income and income mobility varies with household income.

5 Income convergence in South Africa between 2008 and 2012

This approach is now applied to the three waves of the South African NIDS panel dataset¹³. It is a large, nationally representative dataset. The three waves were collected in 2008, 2010 and 2012, respectively. In order to circumvent unbalanced panel issues, we only use the households that were captured and had the same household head in all three waves. That means household members that left the core household were not followed. This provides us with a sample of 2770 households. Table A1 in the appendix contains a list descriptive statistics about the demographic attributes of the households in our sample. Balanced panel weights are used to adjust the sample for attrition across all waves, as explained in Finn and Leibbrandt (2013). This provides us with a sample of 2770 households. Our chosen measure of income, y , is real per capita household income from all sources and with imputations for any missing values. Table A1 in the appendix.

Table 2 reports the estimates for the seven regression coefficients of interest. The regression coefficients in columns 1 and 2 (that correspond to θ_1 and θ_2) are both very close to -0.25, suggesting that approximately 25% of income gaps are eliminated in the two-year periods between surveys. At this rate of convergence we could expect half of the income gap between the richest and poorest South African household to be eliminated every 4.8 years. This seems like a surprisingly high degree of income mobility, but is consistent with the estimate obtained by Fields et al. (2003a) for South Africa using the KIDS panel, which implies that 56% of income gaps should be eliminated over 5 years. The estimates from Fields et al. (2003a) imply an even more rapid speed of converge for other countries: Indonesia (53% over 4 years), Spain (52% over 1 year) and Venezuela (64% over 1 year). Of course, all of these estimates are vulnerable to the presence of measurement error. Note also that the similarity of these two regression coefficients suggest that, as assumed above, the speed income convergence did not change substantially between the two periods considered.

In order to investigate the internal consistency of the estimates reported in Table 2, the same estimates are replicated in the first row of Table 3 along

¹³See Finn and Leibbrandt (2013) for a detailed description of the data.

with (in the second row) the expected values of these coefficients if the estimated regression coefficient θ_1 represents the true convergence parameter β and income is measured without error. Apart from θ_2 , none of the regression coefficients is near its predicted value. The effect of wave 1 income on income growth between waves 2 and 3 (represented by θ_3), and total income growth between waves 1 and 3 (represented by θ_4) are both much smaller than we would have expected given the rapid income growth that occurred between waves 1 and 2, and waves 2 and 3. The coefficient estimates obtained from regressing Δy_3 on y_1 and y_2 (θ_5 and θ_6) are also very different than what we would expect in the absence of measurement error. Instead of values close to zero and θ_1 , we observe estimates that are significantly positive and significantly more negative than θ_1 . Finally, θ_7 reveals a stronger negative correlation between Δy_3 and Δy_2 than we can explain without measurement error.

Another way of testing for the existence of measurement error is to calculate the values of β implied by each regression coefficient estimate. The implied parameter values (reported in the fourth row of Table 3) vary from very small (-0.045) to very large (-0.495), and seem unlikely to have been produced by the same convergence parameter in the absence of measurement error. Our informal method of investigating whether income is measured without error therefore provides evidence against this hypothesis, although it cannot be formally tested in this way.

Equation [6] showed how we can obtain point estimates of β and α using the estimates of θ_1 and θ_3 , which produces estimates of $\hat{\beta} = -0.057$ and $\hat{\alpha} = 0.80$. As suggested in section 4.1, these estimates are used to predict the values of the other five regression coefficients according. These predicted values are reported in row three of Table 3. Unlike the predicted values obtained under the no measurement error assumptions, these predictions are all very close to the estimated values in row 1. The values of β that are implied by each of the regression coefficients if $\alpha = 0.8$ are also reported in row five of Table 3. These values are observed to lie within a narrow range between -0.061 and -0.05. Allowing for classical measurement error thus makes it possible to provide an internally consistent explanation of the estimated regression coefficients, while it is impossible to do so while maintaining the assumption that income is measured without error.

Next, we proceed to estimate the model parameters using the system GMM estimator. The results are shown in Table 4. If we place no restriction on the value of α , then β is estimated to be -0.059. This is similar to the point estimate obtained using equation [6] and implies that only about 6% of income gaps are expected to be eliminated during the two years between survey waves. This is much lower than the estimates obtained by either regressing Δy_2 on y_1 or Δy_3 on y_2 . In fact, this estimate suggests that the conventional approach over-estimates South African income mobility by a factor of between 4 and 5. The implied expected half-life of any income gap is now approximately 27 years, not 5, which would mean that South African households have considerably less economic mobility than previous studies may have led us to believe. At the

same time, it still suggests that there is significant income mobility in South Africa.

The GMM estimate of α indicates that this discrepancy arises because only 80% of the variation in log household income is due to variation in actual incomes, whereas the remaining 20% is due to measurement error. The implied reliability statistic of 0.8 suggests that the NIDS household income measure is more or less as reliable as U.S. self-reported earnings data (Abowd & Stinson, 2013) and more reliable than self-reported earnings data for developing countries (Akee, 2011). It may seem surprising that the convergence parameter can be over-estimated by a factor of nearly five when only a relatively small share of income variation is due to measurement error. However, this bias factor, which can be expressed as $\frac{E(\hat{\theta}_1|\beta,\alpha)}{\beta} = \alpha + \frac{\alpha-1}{\beta}$, can be very large if true income convergence is very slow, even if income measures are relatively reliable. Intuitively, if there is very little inclination for actual incomes to converge towards the mean of the distribution, then most of the observed income convergence will be due to the mean-regressive effect of classical measurement error.

Apart from allowing us to simultaneously estimate the income convergence and data reliability parameters, the GMM estimator has the added advantage of providing estimates that are highly efficient, as can be observed by their small standard errors. Furthermore, it also allows us to formally test the validity of the over-identifying restrictions. The J-test indicates that the GMM estimates can explain all five linearly independent regression coefficients in a way that is internally consistent.

Column 2 of Table 4 reports the GMM estimates of the convergence parameter under the assumption of no measurement error (by restricting $\alpha = 1$). The point estimate of β obtained from using all the regression coefficients is much smaller than suggested by the estimates of either θ_1 or θ_2

on its own. However, the associated J-test also strongly rejects that validity of the associated over-identifying restrictions, which confirms that the assumption of no measurement error is inconsistent with the observed covariance pattern in the data.

As mentioned in section 4.3, a different GMM estimator can be obtained by expressing the variances and covariances of (y_1, y_2, y_3) as functions of $(\beta, \sigma_u^2, \sigma_e^2)$. The estimates from this model are presented in Column 3 of Table 4. The results for the convergence parameter estimate and over-identifying restriction test are both very similar to what is estimated using the original GMM estimator, as is the implied estimate of $\alpha = \frac{\sigma_e^2}{\frac{\sigma_u^2}{1-\beta^2} + \sigma_e^2}$ of 0.801. This similarity is not surprising, since the estimates are based on the same identifying assumptions. The estimates of the variances of the unobservable components are useful because they allow us to investigate the effect of classical measurement error on Shorrocks' index of income mobility.

The estimated two-period mobility index for waves 1 and 2 is $\hat{M}_2 = 0.0624$, whereas the three-period mobility measure for waves 1, 2 and 3 is $\hat{M}_3 = 0.0833$. As one would expect, these estimates indicate positive mobility that increases

along with the accounting period. However, based on what was observed for the estimates of the income convergence coefficient, we would expect measurement error to overstate Shorrocks's index of income mobility, particularly for smaller values of T . The conventional two-wave estimate of $\hat{\beta} = -0.25$ implies that we should expect values of \hat{M}_2 and \hat{M}_3 of around 0.0646 and 0.11, respectively. As with our analysis on micro growth regressions, it appears that longer-run mobility estimates are inconsistent with the supposed high degree of short-run mobility, which suggests that both measures may over-state true income mobility as a result of measurement error. Using the estimates from column 3 of Table 4 and the more general equations for Shorrocks's mobility index that allows for measurement error, we obtain predicted values for \hat{M}_2 and \hat{M}_3 of 0.0636 and 0.0912, respectively. Both of these predictions are closer to the observed sample estimates than the values predicted by the model that assumes no measurement error, which provides additional support for our hypothesis of measurement error induced upwardly biased mobility estimates. Furthermore, if we believe that these adjusted estimates offer a more reliable indication of the true data generating process, then our corrected estimates of the Shorrocks mobility indices are actually $\hat{M}_2 = 0.0147$ and $\hat{M}_3 = 0.026$, both of which indicate substantially less economic mobility than obtained with the naïve estimates that ignore measurement error.

Finally, we use local linear regressions to estimate the nonparametric generalisations of regression coefficients θ_1 and θ_3 . These estimates are plotted (against demeaned y_1 on the x-axis) in Figure 1, and used to calculate nonparametric estimates of $\beta(y_1)$ and $\alpha(y_1)$ according to equation [8]. The resulting estimates of the β and α functions are graphed in Figure 2. The income measure reliability statistic varies between 0.6 for low initial income values and 0.95 for high initial incomes. Instead of 20% of the variation in all household incomes being due to measurement error, this share is as high as 40% for poor households and as low as 5% for rich ones. The income convergence parameter varies between -0.14 for poor households and -0.03 for rich households, which reveals that income mobility also depends on initial income. Whereas our parametric estimates indicated that all household could expect 6% of the income gap between itself and other households to be eliminated, the nonparametric estimate reveals that poor households can expect to experience more upward mobility, whereas rich household should experience comparatively little downward mobility on average.

6 Conclusion

This study proposed a new approach that uses three-wave panel data to estimate income mobility when incomes are measured with error. This approach is applied to a three-wave South African panel dataset. Substantively, we find the conventional method over-estimates the extent of income mobility by a factor of between 4 and 5. This result is robust to the choice of income mobility measure, and occurs because about 20% of variation in reported household income is due to measurement error. Nonparametric estimates show that there is relatively

high (upward) income mobility for poor households, but very little (downward) income mobility for rich households, and that the income is much more reliably captured for rich than for poor households. While these estimates suggest much smaller income mobility in South Africa than previously estimated, mobility is nevertheless substantial, offering particular opportunities for upward mobility of poorer households.

References

- [1] Abowd, J. M., & Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*, 95(5), 1451-1467.
- [2] Agüero, J., Carter, M. R., & May, J. (2007). Poverty and inequality in the first decade of South Africa's democracy: What can be learnt from panel data from KwaZulu-Natal? *Journal of African Economies*, 16(5), 782-812.
- [3] Akee, R. (2011). Errors in self-reported earnings: The role of previous earnings volatility and individual characteristics. *Journal of Development Economics*, 96(2), 409-421.
- [4] Antman, F., & McKenzie, D. (2007a). Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. *The journal of development studies*, 43(6), 1057-1083.
- [5] Antman, F., & McKenzie, D. J. (2007b). Earnings Mobility and Measurement Error: A Pseudo-Panel Approach. *Economic Development and Cultural Change*, 56(1), 125-161.
- [6] Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy*, 223-251.
- [7] Barro, R. J., & Sala-i-Martin, X. (2004). *Economic Growth* (2nd ed.). New York: McGraw-Hill.
- [8] Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement Error in Survey Data. In J. J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics*, Vol 5. New York: Elsevier Science B.V.
- [9] Bound, J., & Krueger, A. B. (1991). The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics*, 1-24.
- [10] Cuesta, J., Ñopo, H., & Pizzolitto, G. (2011). Using Pseudo-Panels To Measure Income Mobility In Latin America. *Review of Income and Wealth*, 57(2), 224-246.

- [11] Fields, G. S. (2008a). A brief review of the literature on earnings mobility in developing countries *ILR Working Paper 2-20-2008*. Ithaca: Cornell University.
- [12] Fields, G. S. (2008b). Income mobility. In L. Blume & S. N. Durlauf (Eds.), *The new Palgrave dictionary of economics*. New York, NY: Palgrave Macmillan.
- [13] Fields, G. S., Cichello, P. L., Freije, S., Menéndez, M., & Newhouse, D. (2003a). For richer or for poorer? Evidence from Indonesia, South Africa, Spain, and Venezuela. *The Journal of Economic Inequality*, 1(1), 67-99.
- [14] Fields, G. S., Cichello, P. L., Freije, S., Menéndez, M., & Newhouse, D. (2003b). Household income dynamics: a four-country story. *The journal of development studies*, 40(2), 30-54.
- [15] Fields, G. S., Duval-Hernández, R., Freije, S., & Puerta, M. L. S. (2014). Earnings mobility, inequality, and economic growth in Argentina, Mexico, and Venezuela. *The Journal of Economic Inequality*, 1-26.
- [16] Fields, G. S., & Viollaz, M. (2013). Can the Limitations of Panel Datasets be Overcome by Using Pseudo-Panels to Estimate Income Mobility? *Universidad Cornell-CEDLAS*.
- [17] Finn, A., & Leibbrandt, M. (2013). Mobility and Inequality in the First Three Waves of NIDS: Southern Africa Labour and Development Research Unit, University of Cape Town.
- [18] Glewwe, P. (2012). How much of observed economic mobility is measurement error? IV methods to reduce measurement error bias, with an application to Vietnam. *The World Bank Economic Review*, 26(2), 236-264.
- [19] Gottschalk, P., & Huynh, M. (2010). Are earnings inequality and mobility overstated? The impact of nonclassical measurement error. *The Review of Economics and Statistics*, 92(2), 302-315.
- [20] Heng, Y., Shi, L., & Quheng, D. (2006). Income Mobility in Urban China. *Economic Research Journal*, 10, 002.
- [21] Jäntti, M., & Jenkins, S. P. (2015). Chapter 10 - Income Mobility. In B. A. Anthony & B. François (Eds.), *Handbook of Income Distribution* (Vol. Volume 2, pp. 807-935): Elsevier.
- [22] Jarvis, S., & Jenkins, S. (1998). How much income mobility is there in Britain? *The Economic Journal*, 108(447), 428-443.
- [23] Khor, N., & Pencavel, J. (2006). Income mobility of individuals in China and the United States. *Economics of Transition*, 14(3), 417-458.
- [24] Lechtenfeld, T., & Zoch, A. (2014). Income Convergence in South Africa: Fact or Measurement Error?

- [25] Lee, N. (2009). Measurement error and its impact on estimates of income and consumption dynamics. *Available at SSRN 1299330*.
- [26] Newhouse, D. (2005). The persistence of income shocks: Evidence from rural Indonesia. *Review of development Economics*, 9(3), 415-433.
- [27] Özler, B. (2007). Not separate, not equal: poverty and inequality in post-apartheid South Africa. *Economic Development and Cultural Change*, 55(3), 487-529.
- [28] Quah, D. T. (1996). Empirics for economic growth and convergence. *European Economic Review*, 40(6), 1353-1375. doi: [http://dx.doi.org/10.1016/0014-2921\(95\)00051-8](http://dx.doi.org/10.1016/0014-2921(95)00051-8)
- [29] Van der Berg, S., Burger, R., Burger, R., Louw, M., & Yu, D. (2005). *Trends in poverty and inequality since the political transition*. Stellenbosch Working Paper 1/2005. Stellenbosch University, Department of Economics.
- [30] Woolard, I., & Klasen, S. (2005). Determinants of income mobility and household poverty dynamics in South Africa. *Journal of Development Studies*, 41(5), 865-897.

Table 1: Regression coefficients and population moments

Parameter		Population mean	
		No measurement error	Classical measurement error
θ_1	$L(y_2 - y_1 y_1) = \theta_1 y_1$	β	$(\beta + 1)\alpha - 1$
θ_2	$L(y_3 - y_2 y_2) = \theta_2 y_2$	β	$(\beta + 1)\alpha - 1$
θ_3	$L(y_3 - y_2 y_1) = \theta_3 y_1$	$\beta(\beta + 1)$	$\alpha\beta(\beta + 1)$
θ_4	$L(y_3 - y_1 y_1) = \theta_4 y_1$	$\beta(\beta + 2)$	$\alpha(\beta + 1)^2 - 1$
θ_5	$L(y_3 - y_2 y_1, y_2) = \theta_5 y_1 + \theta_6 y_2$	0	$\frac{(\beta + 1)^2(\alpha - 1)\alpha}{\alpha^2(\beta + 1)^2 - 1}$
θ_6	$L(y_3 - y_2 y_1, y_2) = \theta_5 y_1 + \theta_6 y_2$	β	$\frac{1 - \alpha(\beta + 1) + \alpha^2\beta(\beta + 1)^2}{\alpha^2(\beta + 1)^2 - 1}$
θ_7	$L(y_3 - y_2 y_2 - y_1) = \theta_7(y_2 - y_1)$	$\frac{1}{2}\beta$	$-\frac{1 - \alpha + \alpha\beta^2}{2(1 - \alpha - \alpha\beta)}$

Table 2: Regression coefficients for South African income regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	Δy_2	Δy_3	Δy_3	$y_3 - y_1$	Δy_3	Δy_3
y_1	-0.249*** (0.0251)		-0.0427** (0.0196)	-0.292*** (0.0254)	0.329*** (0.0295)	
y_2		-0.243*** (0.0227)			-0.495*** (0.0267)	
Δy_2						-0.409*** (0.0280)
Constant	1.825*** (0.174)	1.911*** (0.156)	0.471*** (0.139)	2.296*** (0.176)	1.375*** (0.134)	0.189*** (0.0211)
Observations	2,770	2,770	2,770	2,770	2,770	2,770
R-squared	0.129	0.141	0.004	0.170	0.252	0.194

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3: Regression coefficients and implied parameter values

θ_k	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
Estimated values of θ_k	-0.249*** (0.0251)	-0.243*** (0.0227)	-0.0427** (0.0196)	-0.292*** (0.0254)	0.329*** (0.0295)	-0.495*** (0.0267)	-0.409*** (0.0280)
Predicted values of θ_k if $\beta = -0.249$ and $\alpha = 1$	-0.249	-0.249	-0.187	-0.436	0.000	-0.249	-0.125
Predicted values of θ_k if $\beta = -0.057$ and $\alpha = 0.8$	-0.249	-0.249	-0.043	-0.292	0.330	-0.497	-0.414
Value of β implied by θ_k (if $\alpha = 1$)	-0.249	-0.243	-0.045	-0.159	NA	-0.495	-0.818
Value of β implied by θ_k (if $\alpha = 0.8$)	-0.057	-0.050	-0.057	-0.057	-0.058	-0.060	-0.061

Table 4: GMM estimates for South African income dynamics

	(1)	(2)	(3)
β	-0.0590*** (0.0174)	-0.0886*** (0.00455)	-0.0585*** (0.0174)
α	0.801*** (0.0195)	1 .	
σ_u^2			0.151*** (0.0422)
σ_e^2			0.331*** (0.0304)
Observations	2,770	2,770	2,770
J-test statistic	0.249	73.2	0.251
p-value	0.969	0	0.882

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Local linear regression estimates of $\theta_1(y_1)$ and $\theta_4(y_1)$

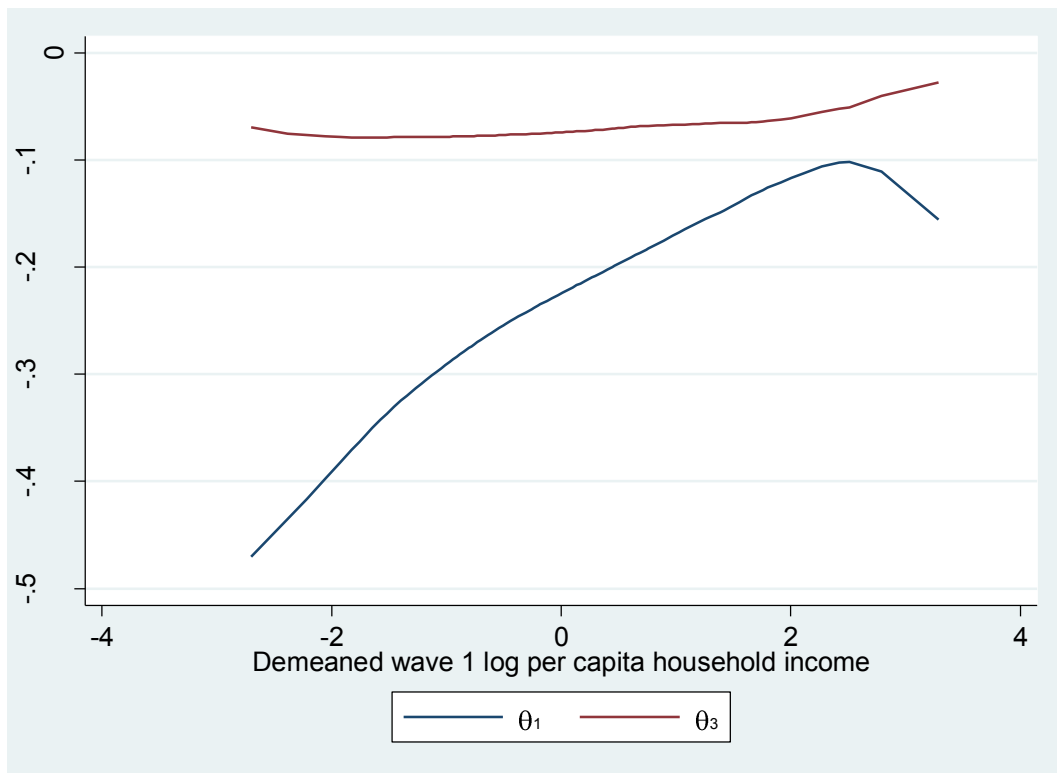
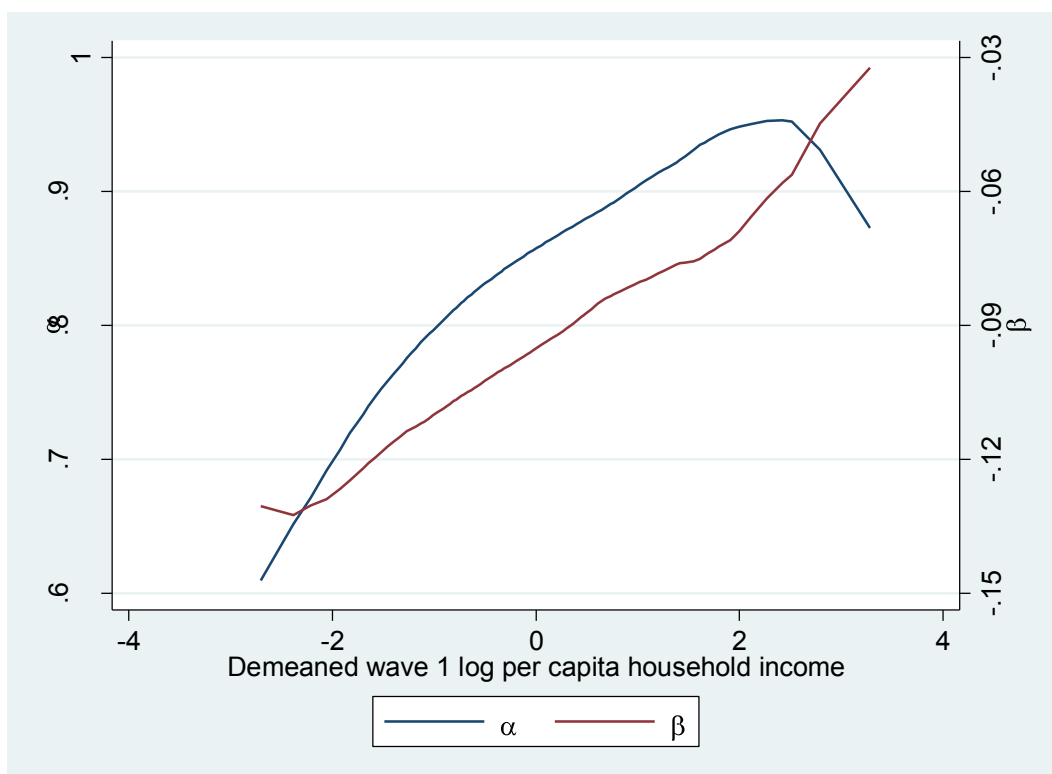


Figure 2: Nonparametric estimates of $\beta(y_1)$ and $\alpha(y_1)$



APPENDIX

Table A1: Demographic attributes of sample (mean values)

	2008	2010	2012
Age (of household head)	49.6	52.1	53.8
Household members	3.8	4.0	4.0
Number of children	1.3	1.3	1.3
Years of schooling (household head)	6.6	6.7	6.8
Share (household head): Black	82%	82%	82%
White	4%	4%	4%
Coloured	12%	12%	12%
Indian	1%	1%	1%
Resides in rural area	49%	49%	48%