Stanford
Center for International Development

Working Paper No. 254

Earnings Mobility and Measurement Error: A Pseudo-Panel Approach

by

Francisca Antman[*] and David J. Mckenzie[**]

September 2005

Stanford University
John A. and Cynthia Fry Gunn Building
366 Galvez Street | Stanford, CA | 94305-6015

[*] Department of Economics, Stanford University
[**] Development Research Group, The World Bank

Earnings Mobility and Measurement Error: A Pseudo-Panel Approach*

Francisca Antman
Department of Economics, *Stanford University*

David J. Mckenzie†
Development Research Group, *The World Bank*

August 29, 2005.

Abstract

The degree of mobility in incomes is often seen as an important measure of the equality of opportunity in a society and of the flexibility and freedom of its labor market. However, estimation of mobility using panel data is biased by the presence of measurement error and non-random attrition from the panel. This paper shows that dynamic pseudo-panel methods can be used to consistently estimate measures of absolute and conditional mobility in the presence of non-classical measurement errors. These methods are applied to data on earnings from a Mexican quarterly rotating panel. Absolute mobility in earnings is found to be very low in Mexico, suggesting that the high level of inequality found in the cross-section will persist over time. However, the paper finds conditional mobility to be high, so that households are able to recover quickly from earnings shocks. These findings suggest a role for policies which address underlying inequalities in earnings opportunities.

JEL classification: O12, D31, C81
Keywords: income mobility, dynamic pseudo panel, measurement error.

*We thank the Stanford Center for International Development (SCID) for research funding; and Aart Kraay and participants at the PACDEC conference for helpful comments.
†Email: dmckenzie@worldbank.org

1 Introduction

The degree of mobility in income is often seen as a measure of the equality of opportunity in a society, and of the flexibility and freedom of movement in the labor market (Atkinson, Bourguignon and Morrisson, 1992). Greater mobility makes the distribution of lifetime incomes more equal for a given level of single period income inequality. On the other hand, Jarvis and Jenkins (1998) note that too much mobility may represent income fluctuations and economic insecurity. Gottschalk and Spolaore (2002) formalize this trade-off in a model with both aversion to inequality and aversion to unpredictability of incomes, finding the socially desirable level of mobility will be less than a level at which there is full reversal of ranks over time. Nevertheless, in many developing countries, the concern is more likely to be that there is too little, rather than too much, mobility. In particular, Piketty (2000) surveys recent theoretical work which finds that the presence of credit constraints can give rise to the possibility of "low-mobility traps", whereby households who need to borrow to finance investment can take a long time to build up wealth.

Measurement of the degree of mobility using panel data on earnings is complicated by the presence of measurement error, and by non-random attrition from the panel. A simple measure of mobility is the slope coefficient from a regression of current period earnings on lagged earnings (e.g. Jarvis and Jenkins, 1998; Fields et al. (2003)). Classical measurement error causes the well-known attenuation bias towards zero in the estimated slope coefficient, leading one to overstate the degree of mobility. The existing literature has attempted to overcome the measurement error problem through the use of instrumental variable methods.[1] Instruments for lagged income have included lagged expenditure (e.g. McCulloch and Baulch, 2000), subjective measures of well-being (Luttmer,

[1] An alternative method used in developed countries has been the use of administrative data from payroll records, where measurement error is likely to be much less (e.g. Aaberge et al., 2002). Such data is much less common in developing countries, and does not allow analysis on the dynamics of earnings of the self-employed, or of total household income.

2002)[2], assets and land holdings (e.g. Fields et al. (2003), Strauss et al. (2004)), and weight (Glewwe and Nguyen, 2002).

A key condition for the validity of such instruments is that any measurement error in the instrument is uncorrelated with the measurement error in income. Glewwe and Nguyen (2002) question this assumption in the case of expenditure, where individuals may systematically underreport both income or expenditure, or interviewers may reduce the level of detailed questioning on both measures. However, even when the instrument is uncorrelated with the measurement error, consistency of the instrumental variables estimator still requires that the instrument also be uncorrelated with the non-measurement error component of the error term in the data generating process. Glewwe and Nguyen state that such an assumption is extremely unlikely to hold when variables such as education or land, which have a causal relationship with income, are used as instruments. In this paper we show further that if the instrument itself follows an AR(1) process, then the instrumental variables estimator will converge to the autocorrelation coefficient of the instrument. As a result, instrumental variables will only be consistent if the instrument is uncorrelated with the measurement error and has the same amount of mobility as earnings. This condition appears extremely unlikely to hold in practice.

The literature has devoted less attention to assessing the impact of attrition on estimates of mobility. However, the typical labor force panel in developing countries reinterviews dwelling units, rather than households, so that households that move attrit from the sample. The Mexican Urban Labor Force Survey (ENEU) used in this study is a quarterly rotating panel which follows this approach, and on average loses 35 percent of the sample due to attrition over the five periods. Thomas, Frankenberg and Smith (2001) discuss the experience of the Indonesia Family Life Survey, which explicitly tracked movers, and do find that those who move are different in terms of initial characteristics than those who stay. Although they do not examine whether changes in income or

[2]Luttmer actually examines mobility in expenditure, rather than income, and uses income and subjective well-being as instruments for expenditure.

other economic conditions are associated with households being more likely to move, one would expect greater geographic mobility to be associated with more income mobility: households experiencing large positive shocks may move to better housing while households experienced large negative shocks may migrate or move to cheaper housing. As a result, the attrition bias will lead panel studies to understate mobility.

This paper shows how dynamic pseudo-panel methods can be used to consistently estimate the degree of income mobility when earnings contain non-classical measurement error. A pseudo-panel tracks cohorts of individuals over repeated cross-sectional surveys (Deaton, 1985). Construction of a pseudo-panel involves taking cohort means within each time period, and this averaging process eliminates individual-level measurement error. Since each household is only observed once, non-random attrition becomes much less of an issue. A further advantage is that repeated cross-sectional surveys are available in more countries and typically over longer time-periods than genuine panels. This allows one to estimate mobility measures over many more time periods than typically used in the panel literature. Gottschalk (1997) notes that many movements in income are transitory, so that individuals who experience an increase in earnings in one year will tend to have a fall in income a few years later. Therefore mobility over several periods may be different from what one would predict based on extrapolating measures based on a one year interval.

This paper uses 58 quarters of household earnings data in Mexico over the period 1987 to 2001 to examine earnings mobility. Mexico's income distribution displays a high degree of cross-sectional inequality, and therefore a high degree of income mobility is of importance in lowering inequality in lifetime distributions of income. However, our pseudo-panel results find very low levels of absolute mobility in Mexico. While OLS estimation would suggest that 33 percent of the gap in income between two randomly selected households would close within one quarter, pseudo-panel analysis finds only 1.2 percent of this gap would be eliminated within a quarter, and only five to seven percent of income differences disappear after five years. The OLS bias appears almost entirely

due to measurement error and does not appear to be much offset by differential attrition of the more mobile. In contrast, while absolute mobility remains low, conditional mobility, defined as the movement in income around a household's fixed effect, is found to be quite rapid. Households which experience bad luck or shocks to labor earnings which take them below the level of income determined by their individual attributes recover almost fully to their expected level within two years. These findings of slow absolute mobility and rapid conditional mobility continue to hold using full income and expenditure from an alternate dataset. As a result, the high levels of inequality seen in a given cross-section are likely to persist over time.

The remainder of the paper is structured as follows. Section 2 discusses estimation of mobility by OLS and IV in the presence of non-classical measurement errors. Section 3 shows how pseudo-panel estimation can allow for consistent estimation. Section 4 describes the data. Section 5 contains the main results of the paper while Section 6 provides an interpretation of the findings. Section 7 concludes.

2 Mobility and Measurement Error

While there are many potential measures of mobility (see Atkinson et al. (1992)), we investigate one of the simplest measures, which is the slope coefficient in a regression of income on its lagged value. This measure is common in much of the empirical literature (e.g. Jarvis and Jenkins, 1998; Fields et al. (2003); Strauss et al. (2004)). Moreover, because this measure is based on a regression framework, pseudo-panel methods and instrumental variables can be applied to deal with measurement error. Our application will investigate mobility in labor income, but the methods which follow can easily be applied to expenditure or other socioeconomic variables.

Consider the data generating process for the actual log income, $Y^*_{i,t}$ of indi-

vidual i at time period t:

$$Y^*_{i,t} = \alpha + \beta Y^*_{i,t-1} + u_{i,t} \tag{1}$$

The coefficient β is a measure of (im)mobility. A value of β of unity indicates that incomes move in step, with no convergence of incomes. If β is greater than unity, there is divergence, and β less than one indicates some convergence of incomes. Gottschalk and Spolaore (2002) consider two aspects of economic mobility. 'Origin independence' measures the degree to which future incomes do not depend on present income. β equal to zero combined with no individual fixed effects in the error term, $u_{i,t}$ would indicate full original independence. They also consider a second aspect, 'reversal', which is the degree to which ranks are reversed over time. A value of β less than zero would indicate some reversal, with individuals with above mean income experiencing a fall in income and poorer individuals getting richer. The socially optimal level of β will involve a trade-off between the degree of aversion to inequality (which favors lower values of β) and the degree of aversion to unpredictability of income (which favors values of β closer to one). Consistent measurement of β is needed to assess the degree of mobility.

However, in practice one observed data which are measured with error. One thus observes:

$$Y_{i,t} = Y^*_{i,t} + \varepsilon_{i,t} \tag{2}$$

We wish to place relatively weak assumptions on the measurement error $\varepsilon_{i,t}$. In particular, Bound and Krueger (1991) compare the Current Population Survey to Social Security Administrative records in the United States and find that the measurement error in earnings is positively autocorrelated over two years, and is negatively correlated with true earnings. These findings violate standard classical measurement error assumptions, and since similar validation studies are not available for developing countries, we wish to allow generally for the possibility of autocorrelation and correlation with true earnings.

Substituting (2) into (1) gives the equation to be estimated in terms of

observed income:

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + \eta_{i,t} \quad (3)$$

$$\text{where } \eta_{i,t} = u_{i,t} + \varepsilon_{i,t} - \beta\varepsilon_{i,t-1} \quad (4)$$

Consider the OLS estimator of β based on equation (3):

$$\widehat{\beta}_{OLS} = \frac{\sum_{i=1}^{N} Y_{i,t} y_{i,t-1}}{\sum_{i=1}^{N} Y_{i,t-1} y_{i,t-1}}$$

where $y_{i,t-1} = Y_{i,t-1} - (1/N)\sum_{i=1}^{N} Y_{i,t-1}$. One can then show under standard assumptions that as the number of observations in the cross-section, N, goes to infinity,

$$\begin{aligned} & \widehat{\beta}_{OLS} \xrightarrow{p} \beta + \theta_{OLS} \\ \text{where } \theta_{OLS} = & \left[E\left(u_{i,t}, Y_{i,t-1}\right) + Cov\left(\varepsilon_{i,t}, \varepsilon_{i,t-1}\right) + Cov\left(\varepsilon_{i,t}, Y^{*}_{i,t-1}\right)\right. \\ & \left. -\beta Var\left(\varepsilon_{i,t-1}\right) - \beta Cov\left(Y^{*}_{i,t-1}, \varepsilon_{i,t-1}\right)\right] / Var\left(Y_{i,t-1}\right) \end{aligned}$$

The term θ_{OLS} is the asymptotic bias and shows that OLS will be inconsistent in general. This inconsistency arises due to the following terms:

i) $E\left(u_{i,t}, Y_{i,t-1}\right)$, the covariance between the current period shock to earnings and last periods measured earnings. The standard concern here is the present of individual fixed effects in the error term $u_{i,t}$, which will lead to this term being positive. This term will also not be zero if earnings shocks, $u_{i,t}$ are autocorrelated.

ii) $Cov\left(\varepsilon_{i,t}, \varepsilon_{i,t-1}\right)$, the covariance between this period's and last period's measurement error terms will be non-zero if measurement errors are autocorrelated. Based on the U.S. validation studies, we would expect this term to be positive.

iii) $Cov\left(Y^{*}_{i,t-1}, \varepsilon_{i,t-1}\right)$, the covariance between the measurement error and true earnings. The results of Bound and Krueger (1991) suggest this term

will be negative. In addition, if the measurement errors are positively autocorrelated, then the covariance between last period's true earnings and the current period's measurement error, $Cov\left(\varepsilon_{i,t}, Y^*_{i,t-1}\right)$, may also be negative.

iv) $Var\left(\varepsilon_{i,t-1}\right)$, the variance of the measurement error. If there are no fixed effects, and the measurement error is classical, then we have:

$$\widehat{\beta}_{OLS} \xrightarrow{p} \beta\left[1 - \frac{Var\left(\varepsilon_{i,t-1}\right)}{Var\left(Y_{i,t-1}\right)}\right] \tag{5}$$

This is the classic attenuation bias towards zero, and would lead one to conclude there is more mobility in income than there actually is.

2.1 Instrumental Variables

In recognition of the effect of measurement error on mobility estimates, several authors have attempted to use instrumental variables methods. As discussed in the introduction, instruments used for income have included education, expenditure, asset holdings, and weight. Let $Z_{i,t-1}$ be the instrument. Then it is assumed that the actual data are related to the instrument according to:

$$Y^*_{i,t-1} = \phi + \gamma Z_{i,t-1} + v_{i,t-1} \tag{6}$$

Where $\gamma \neq 0$ is a necessary condition for instrument relevance. Writing this in terms of the observed $Y_{i,t-1}$ then gives the first-stage equation:

$$Y_{i,t-1} = \phi + \gamma Z_{i,t-1} + v_{i,t-1} + \varepsilon_{i,t-1} \tag{7}$$

Let $z_{i,t-1} = Z_{i,t-1} - (1/N)\sum_{i=1}^{N} Z_{i,t-1}$ denote the demeaned $Z_{i,t-1}$. The instrumental variables estimator of β based on equation (7) being used as a first-stage

for $Y_{i,t-1}$ in equation (3) is then:

$$\begin{aligned}\widehat{\beta}_{IV} &= \frac{\sum_{i=1}^{N} Y_{i,t} z_{i,t-1}}{\sum_{i=1}^{N} Y_{i,t-1} z_{i,t-1}} \\ &= \beta + \frac{\sum_{i=1}^{N} \left(u_{i,t} + \varepsilon_{i,t} - \beta\varepsilon_{i,t-1}\right) z_{i,t-1}}{\sum_{i=1}^{N} Y_{i,t-1} z_{i,t-1}} \end{aligned} \tag{8}$$

In order to determine the probability limit of this estimator, we need to impose some structure on the time series properties of the instrument. Let us assume that:

$$Z_{i,t} = \mu + \rho Z_{i,t-1} + \omega_{i,t} \tag{9}$$

This formulation allows us to vary the degree of autocorrelation in the instrument by varying ρ, and to also consider the case of time invariant instruments such as education, for which $\rho = 0$ and $\omega_{i,t} = \omega_i$. Appendix 1 then shows that as $N \to \infty$,

$$\widehat{\beta}_{IV} \xrightarrow{p} \beta + \frac{\gamma\left(\rho - \beta\right) Var\left(Z_{i,t-1}\right) + E\left(\varepsilon_{i,t} Z_{i,t-1}\right) - \beta E\left(\varepsilon_{i,t-1} Z_{i,t-1}\right) + \lambda}{\gamma Var\left(Z_{i,t-1}\right) + E\left(Z_{i,t-1}\varepsilon_{i,t-1}\right) + E\left(Z_{i,t-1} v_{i,t-1}\right)} \tag{10}$$

where

$$\lambda = \gamma E\left(\omega_{i,t} Z_{i,t-1}\right) + E\left(v_{i,t} Z_{i,t-1}\right) - \beta E\left(v_{i,t-1} Z_{i,t-1}\right) \tag{11}$$

Equation (10) thus shows that consistency of the instrumental variables estimator requires that all of the following conditions hold:[3]

1. The instrument $Z_{i,t-1}$ is uncorrelated with both the current and lagged measurement errors. This appears unlikely to hold when using expenditure as an instrument for income, but appears plausible for measures such as education and body weight.

[3] Of course it is theoretically possible that the bias terms could cancel one another out, so that we could obtain consistency without the separate bias terms all being zero, but this appears unlikely in practice.

2. $\lambda = 0$. This requires that the instrument, $Z_{i,t-1}$ be uncorrelated with the error terms $\omega_{i,t}$, $v_{i,t}$ and $v_{i,t-1}$. This condition will be violated if the true data, $Y^*_{i,t}$ contain individual fixed effects which are correlated with the instrument, or if the dynamic process governing the evolution of the instrument itself contains an individual fixed effect. Again, this restriction appears problematic when using expenditure as an instrument for income, since one might expect individual fixed effects in income and expenditure to be correlated.

3. The degree of autocorrelation in the instrument must perfectly match the degree of autocorrelation in income, that is, $\rho = \beta$. This condition is unlikely to be met by many of the instruments used in the literature. In particular, there is no reason to expect the degree of autocorrelation in asset holdings or in body weight to be the same as in earnings. Note that if conditions 1 and 2 hold, then

$$\widehat{\beta}_{IV} \xrightarrow{p} \rho \tag{12}$$

That is, the instrumental variables estimator will converge to the autocorrelation coefficient in the instrument. Hence, if one uses an instrument which does not vary over time, such as education of adults, then $\rho = 1$, and $\widehat{\beta}_{IV}$ will converge to unity.[4] If one uses an instrument which is white noise, then $\widehat{\beta}_{IV}$ will converge to zero.

These three conditions are unlikely to be met simultaneously by most of the candidate instruments used thusfar in the literature. Instruments such as repeated measures of income are most likely to display the same degree of autocorrelation as true earnings, but also therefore likely to have correlated measurement errors and also potentially have individual fixed effects correlated with those in genuine earnings. Instruments such as body weight, education, and land

[4] Glewwe and Nyugen (2002) also show that the correlation coefficient between current and lagged income will be unity in their IV method when using an instrument which does not vary over time.

holdings are less likely to have measurement errors correlated with the measurement error in earnings, but also be less likely to display identical dynamics to income. As a result, the above analysis suggests that all such IV estimators will deliver inconsistent estimates of mobility.

2.2 Instrumental Variables with Individual Effects

It is common practice in dynamic panel data estimation to worry about the presence of individual fixed effects. As seen above, even when there is no measurement error, the presence of individual fixed effects can result in inconsistent estimates of β from both OLS and from certain instrumental variable estimators. The standard solution is to difference the data and then use further lags of income as an instrument. As our panels are very short, we will follow Arellano (1989) in using $Y_{i,t-2}$ as an instrument for $\Delta Y_{i,t-1}$. The Arellano instrumental variables estimator is then:

$$\beta^A = \frac{\sum_{i=1}^{N} (\Delta Y_{i,t}) Y_{i,t-2}}{\sum_{i=1}^{N} (\Delta Y_{i,t-1}) Y_{i,t-2}} \tag{13}$$

Assume that after removing individual fixed effects, the $u_{i,t}$ are not autocorrelated and are independent of $Y^*_{i,s}$ for $s < t$, and are independent of the measurement error. Then if the measurement error is classical, one can show that as $N \to \infty$,

$$\beta^A \xrightarrow{p} \beta \left(1 - \frac{Var(\varepsilon_{i,t-2})}{(1-\beta) E\left(Y^2_{i,t-2}\right) + \beta Var(\varepsilon_{i,t-2})} \right) \tag{14}$$

Therefore with classical measurement error, the Arellano instrumental variables estimator will be biased towards zero for $0 < \beta < 1$. The presence of measurement error will therefore lead this estimator to overstate the degree of mobility.[5]

[5] Again if we allow for non-classical measurement error the bias term becomes more complicated and theoretically difficult to sign.

3 Pseudo-panel Estimation

We propose using pseudo-panel methods to consistently estimate the degree of income mobility in the presence of measurement error. A pseudo-panel tracks cohorts of individuals, such as birth cohorts, or birth-education cohorts, over repeated cross-sectional surveys. Since a new sample of individuals is taken in each period, the use of a pseudo-panel will also greatly reduce the effect of attrition on mobility estimates. The use of the pseudo-panel will capture mobility which is accompanied by movement within the cross-sectional survey domain. However, it will not capture mobility which arises from migration into or out of the survey area. Moffitt (1993), Collado (1997), McKenzie (2001, 2004) and Verbeek and Vella (forthcoming) discuss conditions under which one can consistently estimate linear dynamic models with pseudo-panels. Our aim here is to show that these methods can also deal with the measurement error problems facing panel data models.

Begin by taking cohort averages of equation (3) over the n_c individuals observed in cohort c at time t :

$$\begin{aligned} \overline{Y}_{c(t),t} &= \alpha + \beta \overline{Y}_{c(t),t-1} + \\ &\quad \overline{u}_{c(t),t} + \overline{\varepsilon}_{c(t),t} - \beta \overline{\varepsilon}_{c(t),t-1} \end{aligned} \tag{15}$$

where $\overline{Y}_{c(t),t} = (1/n_c) \sum_{i=1}^{n_c} Y_{i(t),t}$ denotes the sample mean of Y over the individuals in cohort c observed at time t. With repeated cross-sections, different individuals are observed each time period. As a result, the lagged mean $\overline{Y}_{c(t),t-1}$, representing the mean income in period $t-1$ of the individuals in cohort c observed at time t, is not observed. Therefore we replace the unobserved terms with the sample means over the individuals who are observed at time $t-1$, leading to the following regression for cohorts $c = 1, 2, ..., C$ and time periods $t = 2, ..., T$:

$$\begin{aligned} \overline{Y}_{c(t),t} &= \alpha + \beta \overline{Y}_{c(t-1),t-1} + \\ &\quad \overline{u}_{c(t),t} + \overline{\varepsilon}_{c(t),t} - \beta \overline{\varepsilon}_{c(t),t-1} + \lambda_{c(t),t} \end{aligned} \tag{16}$$

where

$$\lambda_{c(t),t} = \beta \left(\overline{Y}_{c(t),t-1} - \overline{Y}_{c(t-1),t-1} \right)$$

As shown in McKenzie (2004), as the number of individuals in each cohort becomes large, $\lambda_{c(t),t}$ converges to zero, and hence we will ignore this term in what follows. Consider then the mean measurement error in income at time t for individuals in cohort c, $\overline{\varepsilon}_{c(t),t}$. As the number of individuals in the cohort gets large, $n_c \to \infty$, we have that:

$$\overline{\varepsilon}_{c(t),t} = \frac{1}{n_c} \sum_{i=1}^{n_c} \varepsilon_{i(t),t} \xrightarrow{p} E\left(\varepsilon_{i(t),t}\right) = 0$$

The last equality assumes that there is no cohort-level component to measurement errors. We can allow for cohort-specific effects in equation (16), in which case we need only assume that there is no time-varying cohort-level component to measurement errors. This assumption does allow for arbitrary autocorrelation in individual measurement errors over time, and for measurement errors to be correlated with true values, provided that this correlation does not vary at the cohort level over time. Under these assumptions, construction of the pseudo-panel, by averaging over the observations in a cohort, will average out the measurement errors.

As a result, with sufficient observations per cohort, the measurement errors do not affect the consistency of estimates from equation (16). The precise method for estimating equation (16) depends on the assumptions one wishes to make about the individual level shocks to earnings, $u_{i,t}$, and on the dimensions of the pseudo-panel in practice. McKenzie (2004) discusses these choices. In particular, if the $u_{i,t}$ contain individual fixed effects but no time-varying cohort-level component, one can estimate β consistently by OLS on the cohort average equation (16) with the inclusion of cohort dummies. This will be consistent as the number of individuals per cohort gets large. If the individual level shocks to earnings contain a common cohort component, then in addition to a large number of individuals per cohort, one also needs a large number of cohorts or a large number of time periods for consistency. With many cohorts and less

individuals per cohort, instrumental variables methods can be used in which lagged cohort means are used as instruments (see Collado, 1997). In our empirical context we choose cohorts to allow for a large number of individuals per cohort, and therefore can use OLS on the cohort means for estimation.

3.1 Mobility and heterogeneity

The most basic specification is therefore to assume that there are no individual fixed effects, in which case one uses the pseudo-panel to estimate β in the following equation:

$$\overline{Y}_{c(t),t} = \alpha + \beta \overline{Y}_{c(t-1),t-1} + \omega_{c(t),t} \tag{17}$$

If Y is the level of income, then $\beta < 1$ then tells us that a household with income below the mean in period $t-1$ will experience more rapid income growth than richer households. This is known as absolute convergence in the macro growth literature (Barro and Sala-i-Martin,1999).

If the data generating process contains individual fixed effects, one can instead include cohort fixed effects, and estimate β in the following equation:

$$\overline{Y}_{c(t),t} = \alpha_c + \beta \overline{Y}_{c(t-1),t-1} + \omega_{c(t),t} \tag{18}$$

An estimate of β which is less than unity from equation (18) can be interpreted as saying that a household which is below its own mean income grows faster. This is called conditional convergence in the growth literature. Allowing for individual fixed effects greatly increases the speed of convergence across countries. However, as Islam (1995, p. 1162) observes, "by being more successful (through the panel framework) in controlling for further sources of difference in the steady state of income, we have, at the same time, made the observed convergence hollower...There is probably little solace to be derived from finding that countries in the world are converging at a faster rate, when the points to which they are converging remain very different".

An analogous argument can be made in our context of income mobility in household data. Estimation of equation (17) gives us an estimate of 'absolute

mobility', which tells us the extent to which households move around in the overall income distribution. This is the measure that most closely corresponds to the idea that mobility can lower lifetime inequality and provide equality of opportunity. Estimation of equation (18) in contrast can be thought of as giving an estimate of 'conditional mobility', telling us whether households move around relative to their own average income. This relates somewhat to the concept of mobility as a measure of flexibility and efficiency of the labor market. We will provide estimates of mobility under both specifications and discuss further the interpretation of these two measures in Section 6.

4 Data

To investigate earnings mobility in Mexico we use the Encuesta Nacional de Empleo Urbano(ENEU), Mexico's national urban employment survey, conducted by the Instituto Nacional de Estadística, Geografía e Informática (INEGI). The sampling unit is a dwelling or housing structure, and demographic information is collected on the household or households occupying each dwelling. An employment questionnaire is then administered for each individual aged 12 and above in the household, providing detailed information on occupation, labor hours, labor earnings, and employment conditions. The survey is designed as a rotating panel, with households interviewed for five consecutive quarters before exiting the survey. In each new round the household questionnaire records absent members, adds any new members who have joined the household, and records any changes in schooling that have taken place. If none of the original group of household members is found to be living in the dwelling unit in the follow-up survey, the household is recorded as a new household (INEGI, 1998). As in many labor force surveys in developing countries, the interviewers do not track households which move, so any household which moves attrits from the panel.

We use data from the first quarter of 1987 through to the second quarter

of 2001[6], providing 58 quarters of data. Over this period the ENEU expanded coverage from 16 cities in 1987 to 34 cities by the end of 1992 and 44 cities by the second quarter of 2001. We include all 39 cities present by the end of 1994, although our results are robust to restricting the sample to just the 16 cities present in all years.

The ENEU only collects data on labor earnings for each household member in their principal occupation. We add this over household members and deflate by the Consumer Price Index for the relevant quarter from the Bank of Mexico to obtain real household labor earnings. To focus only on households for whom labor earnings are likely to be a main source of income, we restrict our sample to households with heads aged 25 to 49 years old. On average two percent of the observations have household labor income of zero. Using data from the ENIGH income and expenditure survey, which does include non-labor sources of income, we calculate that labor income represents 95 percent of total monetary income for urban households with heads in the 25-49 year old age range. In Section 5.2 we examine how mobility in labor earnings compares to estimated mobility in full income and in expenditure.

For our panel data analysis we then have 54 five-quarter panels, beginning with the panel of 3930 households which were sampled from the first quarter of 1987 through to the first quarter of 1988, and ending with the panel of 11,158 households that were sampled from the second quarter of 2000 through to the second quarter of 2001. We use unbalanced panels. Attrition is comparable to dwelling-based labor force surveys in other developing countries. Ten percent of households are observed for only one quarter, while approximately 65 percent of households can be followed for all five quarters.

We form pseudo-panels based on the household head's year of birth and education level. Cohorts are defined by the interaction of five year birth intervals and three education levels (primary schooling or less, 7 to 12 years education, and more than 12 years education). For example, all household heads born

[6]Since the second quarter of 2001, the ENEU was replaced by the ENE, which has now become the ENET - a national quarterly employment survey.

between 1960 and 1964 with primary schooling or less would form one cohort. The household head is defined as the person recognized as the head by the other household members and is generally male. McKenzie (2003) shows that there is no significant change in who is the head for individuals aged 25 to 49, even during the peso crisis in 1995.

A potential concern with the panel data is that more economically mobile households may move, and so the panel will be a selected sample of less mobile households. In order to ensure that the pseudo-panel does not suffer from the same problem, we construct our pseudo-panel using only the households who are in their first wave of the interview. As a result, we use just over 20 percent of the households available in any given cross-section, since the remaining households are those which are being re-interviewed. We restrict the sample further to cohorts with more than 100 observations in a given wave in order to be able to apply the asymptotic theory developed above which relies on a large number of observations per cohort. Approximately 9 percent of cohort-period observations have fewer than 100 households, and including these additional observations does not qualitatively affect our results. After these restrictions, we are left with a pseudo-panel over 58 quarters with 842 cohort-quarter observations.

5 Results

Panel A of Table 1 provides the estimates of the coefficient on quarterly lagged log income from a variety of different estimation methods. Column 1 provides the panel data OLS estimate, 0.668, which is significantly less than unity and suggests substantial mobility within a quarterly period. Adding cohort dummy variables in Column 2 lowers the coefficient estimate further to 0.598. Columns 3 and 4 provide the panel data instrumental variables estimates. As a labor force survey, the ENEU contains few of the variables commonly used in the literature as instruments. We use the education of the household head, and an asset index constructed as the first principal component from questions on the

household dwelling unit[7]. Both of these variables are highly autocorrelated over time, and in accordance the result in equation (12), we obtain an estimate of β very close to unity, 0.99. In contrast, when we employ the second lag of log income as an instrument and employ the Arellano (1992) estimation method, the estimate of β is -0.062, which would indicate full origin independence and in fact some slight reversal in income. This accords with our theoretical result that this estimate will be biased towards zero.

Columns 5 and 6 provide our pseudo-panel estimates of β. When we do not allow for individual effects through cohort-specific intercepts, the estimate of β is 0.988, while after allowing for individual effects we obtain an estimate of β of 0.832. Comparing these results with those in Columns 1 through 4, we see that the OLS estimates greatly overstate mobility, as does the Arellano estimate. The IV estimate using instruments which are strongly autocorrelated happens to give results similar to the pseudo-panel estimate for absolute mobility. This is a consequence of mobility being low over this quarterly period: as equation (12) showed, we would expect to get a coefficient of 0.99 from the IV estimation here regardless of the level of mobility in income, since education and the asset index do not vary much from one period to the next.

Approximately two percent of our households have zero labor income in a given period, and are omitted when calculating log income. In Panel B of Table 1 we therefore repeat the analysis using the level of income, which allows us to include these zeros. The results are qualitatively very similar to those in Panel A, suggesting that the exclusion of these few zero observations does not make a substantive difference.

The use of pseudo-panel analysis allows us to examine mobility over longer time periods than would be possible with the five quarter genuine panels avail-

[7]The asset index is the first principal component of a series of questions about the characteristics of the dwelling unit (type of floor, materials used in the roof and walls, total rooms, whether the household has a separate kitchen, and access to electricity, sewerage, water and telephone). These questions have only been asked since the third quarter of 1994, and are only asked once a year, so by assumption are perfectly autocorrelated within the year.

able in Mexico. Table 2 provides estimates of the mobility coefficient over one quarter, one year, two year, and five year time periods. Since not all cohorts are aged between 25 and 49 in every quarter, less cohort-period observations are available for longer intervals. Table 2 presents results from the balanced pseudo-panel, where the same cohort-quarter observations are used for estimation over different time lags.[8] Columns 1 through 4 provide the estimates of absolute mobility, while Columns 5 through 8 include cohort fixed effects and therefore give measures of conditional mobility. Absolute mobility increases slightly as one increases the time frame, but the estimate of β is still 0.933 over two year intervals and 0.950 over five year intervals.[9] Thus while poorer households experience slightly faster income growth than richer households, a household which has 10 percent higher income than another household today is estimated to still have 9.5 percent higher income five years later.

In contrast, Table 2 shows a high degree of conditional mobility. A ten percent difference in income between two households with the same fixed effect is reduced to a 8.3 percent difference after one quarter, a 5.5 percent difference after one year, and only a 0.5 percent difference after two years. By five years, the households have reversed rankings.

5.1 Mobility and Attrition

Measurement error will result in both OLS and IV methods giving inconsistent measures of mobility. However, a second source of potential bias in mobility estimates based on genuine panel data is that of non-random attrition. This is particularly likely to be a concern in many developing country contexts in which panel surveys track dwelling units, rather than households, over time. Thomas, Frankenberg and Smith (2001) note that this is the standard protocol for follow-up surveys conducted as part of the World Bank's Living Standards Measurement Study, with second round follow-up rates of 87 percent in Cote

[8] The point estimates for the unbalanced pseudo-panel are very similar to those for the balanced pseudo-panel and are available upon request.

[9] The five-year coefficient is not statistically significant from the two-year coefficient.

d'Ivoire, 55 percent in Peru, and 50 percent in Ghana. In the Mexican urban labor force survey used in this paper, 65 percent of households are followed for all five quarters. Thomas et al. (2001) report that the follow-up rate in the second round of the Indonesia Family Life Survey would have been 84 percent instead of 94 percent if they had not tracked households which move. Failure to follow households which move is likely to understate mobility in both the OLS and IV estimates, since it appears likely that households which move dwellings are likely to have experienced greater income changes than households which stay put. Although correction for attrition is possible under certain structural assumptions, most studies of mobility do not attempt to address this issue.[10]

We therefore now investigate how much of the difference between our pseudo-panel estimates and panel data estimates is due to non-random attrition rather than measurement error. We begin by examining the determinants of who attrits. Table 3 presents marginal effects from probit estimation of two types of attrition. Column 1 considers households which attrit after only one round of interviews. These household heads are younger, less likely to be married, have smaller household sizes and larger incomes than household heads who remain for two or more waves of the survey. However, while these differences are significant given the large number of observations, the magnitude of the effects is rather small. In Columns 2 through 5, we look at households which appear in the first two quarters of the survey and examine the determinants of attriting before their full five quarters are completed. This allows us to examine whether attrition is related to the change in income experienced by the household between the first two waves. We find that both the change in income or log income, and the absolute value of this change, are positively associated with subsequent attrition from the panel. However, a one standard deviation change in either the change in income or absolute value of the change in income is associated with less than a 0.01 increase in the probability of attrition. Figure 1 shows that the kernel density of the change in income between periods 1 and 2 is very similar for

[10] An exception is Lokshin and Ravallion (2004), who include a correction for attrition in their study of non-linear income dynamics.

households which attrit to those who do not.

These results suggest that while attrition is more common amongst households which experience greater income mobility, the magnitude of the bias is likely to be rather small. However, a concern might be that households which experience the largest absolute changes in income move houses and attrit out of the survey before the next quarter's survey can be completed. Since these income movements are by assumption unobserved, we can not directly examine them. Instead, in Table 4 we examine how much our pseudo-panel estimates of absolute and conditional mobility differ when we consider only households which don't attrit. We classify households according to whether they participate in all five quarters of the ENEU survey or not, and restrict our analysis to the cohort-quarter observations where we have at least 100 observations per cohort in each group. Column 1 repeats the quarterly pseudo-panel estimate of β in the absolute mobility regression for the full sample. Column 2 creates a pseudo-panel of non-attritors by taking the first wave observations for households which are observed in all five waves. Column 3 creates a pseudo-panel of attritors, by taking the first wave observations for households which are not observed for at least one of the four remaining waves. The estimate of β for the non-attritors pseudo-panel of 0.987 is very close in magnitude and not statistically different from the 0.991 coefficient for the full sample. The attritors pseudo-panel estimate of 0.977 suggests slightly greater absolute mobility among the attriting households[11], but one can not reject equality of the coefficients in the non-attritors and attritors samples. These results therefore suggest that there is very little bias from attrition in estimating mobility with a balanced panel.

Columns 4 through 8 examine conditional mobility of the non-attriting and

[11] Note that households in this pseudo-panel are by definition households that would attrit in the next 4 quarters whenever you sample them. This is a subset of the group of households which happen to attrit in an observed five-quarter period. They are thus households which are likely to have even greater geographic and income mobility than the average attriting household.

attriting households. In Columns 5 and 6 we restrict the cohort effects to be equal for the two samples, while Columns 7 and 8 allow them to differ. Conditional mobility is found to not differ between the two groups when we restrict the cohort effects to be the same for non-attritors and attritors. However, once we allow the cohort effects to vary, the attritors do show somewhat greater conditional mobility than the non-attritors. A 10 percent difference in income between two households with the same fixed effect would be reduced to a 8.2 percent difference after one quarter in the non-attritors sample, and a 7.5 percent difference in the attritors sample.

Overall these results show that attrition has a rather small impact on measurement of mobility, and therefore that measurement error is the main source of bias in the OLS genuine panel estimation. While those who attrit are exhibit slightly more income mobility, the fact that 35 percent of households attrit over the five quarter panel leads us to speculate that changes in income are only one of a large number of reasons why households attrit. A host of idiosyncratic reasons for non-response, temporary absence, and refusal to answer are likely to mitigate the impact of attrition arising from income changes.

5.2 Mobility in Full Income and in Expenditure

The above analysis has been for mobility of household labor earnings in urban Mexico. We can compare mobility in labor earnings with mobility in total household income and in expenditure using Mexico's national income and expenditure survey, the Encuesta Nacional de Ingreso-Gasto de los Hogares (ENIGH). The ENIGH has been carried out in third quarter of the year on a biannual basis since 1992, and we use the six surveys from 1992 to 2002. Each round surveys a new random sample of approximately 10,000 to 14,000 households, so we do not have a panel of households. We can, however, form cohorts based on the same five year birth intervals and three levels of education of the household head as above, and follow cohorts over time. We consider two subsamples of the data. The first consists of urban households, defined as households in areas of

population 100,000 or more, which allows comparison with the ENEU survey. The second is rural households in areas of population of 15,000 or fewer. Out of the 105 cohort-period observations, we have 82 observations in urban areas and only 53 observations in rural areas for which 100 or more households are surveyed within the cohort.

We examine mobility in four different measures of household resources. The first is household income from the primary occupation of each member, which is the measure used in the ENEU. The second, total monetary income, includes all household cash income, including income earned from transfers, pensions, rent, interest, and from non-primary jobs. The third measure, full income, adds non-monetary sources of household income, which includes the value of all home-produced consumption and of any goods received as transfers. The fourth measure is full expenditure, which includes all monetary expenditure and home-produced consumption items. Over the six survey rounds household primary labor earnings has a correlation of 0.91 with total monetary income, 0.83 with full income, and 0.58 with full expenditure.

Panel A of Table 5 presents the estimated slope coefficients from equation (17) for these four measures. For urban households the four measures give very similar levels of absolute mobility. The estimates of β range from 0.86 to 0.89. The rural estimates range from 0.65 (primary wage income) to 0.80 (full expenditure). The point estimates would therefore suggest that there is more absolute mobility in rural areas than in urban areas, and that rural wage income is more mobile than rural expenditure. However the limited number of rural observations results in large standard errors and we can not reject equality of the rural and urban coefficients. The coefficient on log primary wage income for urban households is 0.87 compared to 0.93 for the equivalent measure in the ENEU data. This difference is not statistically significant.

Panel B of Table 5 adds cohort fixed effects and presents the estimated slope coefficients from equation (18). The point estimates suggest very high rates of conditional mobility, with the slope coefficients close to zero. The point estimates also show less conditional mobility in expenditure than in wage income.

The ENIGH data only includes 6 time periods, so with the inclusion of cohort fixed effects, identification of the slope coefficient comes from within-cohort changes in income over this small number of periods. As a result, the standard errors are large, giving wide confidence intervals for conditional mobility. Nevertheless, the coefficient of 0.08 for urban primary wage income is very close to the 0.05 coefficient obtained using the ENEU data.

6 Interpretation

Our results show rather limited absolute mobility in income and expenditure in Mexico, but rapid conditional mobility. In order to interpret this result further, recall the data generating equation for household income at time t given in (1), written to explicitly include the individual fixed effects:

$$Y^*_{i,t} = \alpha_i + \beta Y^*_{i,t-1} + u_{i,t} \tag{19}$$

This can be rewritten as:

$$\begin{aligned} Y^*_{i,t} &= \alpha_i \left(\frac{1-\beta^t}{1-\beta} \right) + \beta^t Y^*_{i,0} \\ &\quad + \left(\sum_{s=0}^{t-1} \beta^s u_{i,t-s} \right) \end{aligned} \tag{20}$$

This partitions current household income into a term due to the household's fixed effect in income growth, a term which represents the effect of initial differences in household income, and a term which represents the cumulative impact of shocks to labor earnings. Comparing the current income of households i and j, we then have that:

$$\begin{aligned} Y^*_{i,t} - Y^*_{j,t} &= (\alpha_i - \alpha_j) \left(\frac{1-\beta^t}{1-\beta} \right) + \beta^t \left(Y^*_{i,0} - Y^*_{j,0} \right) \\ &\quad + \sum_{s=0}^{t-1} \beta^s \left(u_{i,t-s} - u_{j,t-s} \right) \end{aligned} \tag{21}$$

High rates of conditional mobility then imply that if household j has lower current income than household i due to having lower initial income ($Y^*_{j,0} < Y^*_{i,0}$), or a series of bad luck in earnings innovations, household j will rapidly experience faster income growth than household i. However, more rapid conditional mobility only acts to slow the divergence in incomes which comes from differences in fixed effects. When $0 < \beta \leq 1$, $\alpha_i > \alpha_j$ will cause the income gap between household i and j to widen each period, with the rate of expansion greater the larger is β. When $\beta = 0$ (origin independence), initial differences in income and differences in earnings innovations will have no effect on current differences in income, but incomes will always differ by $\alpha_i - \alpha_j$.

Taking cross-sectional variances of equation (20) allows us to see the implications for inequality. We have:

$$\begin{aligned} Var_i\left(Y^*_{i,t}\right) &= Var\left(\alpha_i\right)\left(\frac{1-\beta^t}{1-\beta}\right)^2 + \beta^{2t} Var_i\left(Y^*_{i,0}\right) \\ &\quad + Var_i\left(\sum_{s=0}^{t-1} \beta^s u_{i,t-s}\right) \end{aligned} \tag{22}$$

Cross-sectional inequality in incomes then depends on the degree of inequality in fixed effects, inequality in initial incomes, and inequality in earnings shocks. A higher degree of conditional mobility reduces inequality by lessening inequalities in initial incomes and in earnings shocks, but inequality may still remain high if there is considerable variation in the fixed effects across households.

In terms of the concepts used to motivate the study of mobility, one interpretation is to consider the α_i's as measuring a combination of innate differences in earnings ability and of differences in 'opportunity'. Inequality in the fixed effects therefore would reflect differences in the education and health care of individuals, as well as factors such as discrimination which prevents certain individuals from being able to work in particular occupations. Under this view, β can then be seen as measuring the degree of flexibility and freedom in the labor market. Given predetermined individual attributes, β measures how rapidly individuals who are earning too little or too much relative to their individual abilities and

opportunities regress to their mean level of earnings.

Our finding of slow absolute mobility but rapid conditional mobility has several implications for further study of Mexican income differences. Our finding of rapid conditional mobility suggests that households are able to recover quickly from bad luck and shocks to labor earnings, and therefore that the high level of inequality in Mexican income is not due to income shocks having long-term effects. However, the high rate of conditional mobility coupled with the fact that absolute mobility remains low means that household fixed effects are important and that income differences among households will persist over many years. These fixed effects represent everything specific to a household that has a persistent effect on their income. This includes the education, language, gender, and birth cohort of the household head; household demographic factors; the institutional environment facing a particular household; and other factors that determine labor income such as innate ability, ability to work with others, and entrepreneurial prowess. The challenge for future work is to determine the types of policy interventions which can reduce differences in these fixed effects. Examples may include interventions in health and education and improvements in labor market institutions.

7 Conclusions

We have shown that dynamic pseudo-panel estimation can be used to consistently estimate the degree of earnings mobility, even in the presence of non-classical measurement error. In contrast, OLS and instrumental variables estimators will give biased estimates. Although pseudo-panel estimation also greatly reduces the potential bias from attrition of the most mobile, in practice we find that most of the bias in genuine panel estimation in the Mexican case is due to measurement error, and not attrition.

Our results indicate that overall mobility in earnings, income, and expenditure, is low in Mexico, whereas households are quite mobile around their individual effects. This suggests a role for policy interventions which aim to lower

inequality amongst households in the attributes they bring to the labor market, such as the education and health interventions occurring under the Oportunidades program. In companion work (Antman and McKenzie, 2005), we are investigating whether there are non-linearities in earnings dynamics, which coupled with individual heterogeneity may result in the low levels of mobility we observe being accompanied by poverty traps.

Appendix 1:

Consider:

$$\widehat{\beta}_{IV} = \beta + \frac{\frac{1}{N}\sum_{i=1}^{N}\left(u_{i,t} + \varepsilon_{i,t} - \beta\varepsilon_{i,t-1}\right) z_{i,t-1}}{\frac{1}{N}\sum_{i=1}^{N} Y_{i,t-1} z_{i,t-1}} \tag{23}$$

Let us consider each of the various components of the numerator of the fraction in (23). A standard law of large numbers gives that:

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t} z_{i,t-1} \xrightarrow{p} E\left(\varepsilon_{i,t} z_{i,t-1}\right) \tag{24}$$

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{i,t-1} z_{i,t-1} \xrightarrow{p} E\left(\varepsilon_{i,t-1} z_{i,t-1}\right) \tag{25}$$

Consider next the term $(1/N)\sum_{i=1}^{N} u_{i,t} z_{i,t-1}$. To examine this term, first substitute equation (9) into (7) to get:

$$Y_{i,t} = \phi + \gamma\mu + \gamma\rho Z_{i,t-1} + \gamma\omega_{i,t} + v_{i,t} + \varepsilon_{i,t} \tag{26}$$

Next substitute (7) into (3) to get:

$$Y_{i,t} = \alpha + \beta\phi + \beta\gamma Z_{i,t-1} + \beta v_{i,t-1} + u_{i,t} + \varepsilon_{i,t} \tag{27}$$

Equating equations (27) and (26) then gives:

$$\begin{aligned} u_{i,t} &= (\phi + \gamma\mu - \alpha - \beta\phi) + \gamma(\rho - \beta) Z_{i,t-1} \\ &\quad + \gamma\omega_{i,t} + v_{i,t} - \beta v_{i,t-1} \end{aligned} \tag{28}$$

From (28) we then have:

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} u_{i,t} z_{i,t-1} \\ &\xrightarrow{p} \gamma(\rho - \beta) Var(Z_{i,t-1}) + \lambda \end{aligned} \tag{29}$$

where

$$\lambda = \gamma E(\omega_{i,t} Z_{i,t-1}) + E(v_{i,t} Z_{i,t-1}) - \beta E(v_{i,t-1} Z_{i,t-1}) \tag{30}$$

From (7) we also have that the denominator:

$$\frac{1}{N}\sum_{i=1}^{N} Y_{i,t-1} z_{i,t-1} \xrightarrow{p} \gamma Var(Z_{i,t-1}) + E(Z_{i,t-1}\varepsilon_{i,t-1}) + E(Z_{i,t-1} v_{i,t-1}) \tag{31}$$

Substituting (24), (25), (29) and (31) into (23) gives equation (10).

References:

Aaberge, Rolf, Anders Björkland, Markus Jäntti, Marten Palme, Peder Pedersen, Nina Smith, and Tom Wennemo (2002) "Income Inequality and Income Mobility in the Scandinavian Countries Compared to the United States", *Review of Income and Wealth* 48(4): 443-469.

Antman, Francisca and David J. McKenzie (2005) "Poverty Traps and Nonlinear Income Dynamics with Measurement Error and Individual Heterogeneity", Mimeo. Stanford University.

Arellano, Manuel (1989) "A Note on the Anderson-Hsiao Estimator for Panel Data", *Economics Letters* 31: 337-41.

Arellano, M. and S. Bond (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies* 58: 277-97.

Atkinson, A.B., F. Bourguignon and C. Morrisson (1992) *Empirical Studies of Earnings Mobility, Fundamentals of Pure and Applied Economics* 52, Harwood Academic Publishers: Philadelphia.

Barro, Robert J. and Xavier Sala-i-Martin (1999) *Economic Growth.* The MIT Press: Cambridge, MA.

Baulch, Bob and John Hoddinott (2000) "Economic Mobility and Poverty Dynamics in Developing Countries", *Journal of Development Studies* 36(6): 1-24.

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

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

Fields, Gary S., Paul L. Cichello, Samuel Freije, Marta Menéndez and David Newhouse (2003) "For Richer or for Poorer? Evidence from Indonesia, South Africa, Spain and Venezuela", *Journal of Economic Inequality* 1: 67-99.

Glewwe, Paul and Phong Nguyen (2002) "Economic Mobility in Vietnam in the 1990s", *World Bank Policy Research Working Paper* No. 2838.

Gottschalk, Peter (1997) "Inequality, Income Growth, and Mobility: The Basic Facts", *Journal of Economic Perspectives* 11(2): 21-40.

Gottschalk, Peter and Enrico Spolaore (2002) "On the evaluation of economic mobility", *Review of Economic Studies* 69: 191-208.

Instituto Nacional de Estadística, Geografía e Informática (INEGI) (1998) *Documento Metodológico de la Encuesta Nacional de Empleo Urbano*, INEGI: Aguascalientes, Mexico.

Islam, Nazrul (1995) "Growth Empirics: A Panel Data Approach", *Quarterly Journal of Economics* 110(4): 1127-70.

Jarvis, Sarah and Stephen P. Jenkins (1998) "How much income mobility is there in Britain", *The Economic Journal*, 108: 428-443.

Lokshin, Michael and Martin Ravallion (2004). "Household Income Dynamics in Two Transition Economies", Studies in Nonlinear Dynamics and Econometrics 8(3), Article 4. http://www.bepress.com/snde/vol8/iss3/art4

Luttmer, Erzo F.P. (2002) "Measuring Economic Mobility and Inequality: Disentangling Real Events from Noisy Data", Mimeo. Harris School of Public Policy Studies, University of Chicago.

McCulloch, Neil and Bob Baulch (2000) "Simulating the Impact of Policy Upon Chronic and Transitory Poverty in Rural Pakistan", *Journal of Development Studies* 36(6): 100-130.

McKenzie, David J. (2001) "Estimation of AR(1) models with unequally-spaced pseudo-panels", *Econometrics Journal*, 4: 89- 108

McKenzie, David J. (2003) "How do Households Cope with Aggregate Shocks? Evidence from the Mexican Peso Crisis", *World Development* 31(7): 1179-99.

McKenzie, David J. (2004) "Asymptotic theory for heterogeneous dynamic pseudo-panels", *Journal of Econometrics* 120(2): 235-262.

Moffitt, Robert (1993) "Identification and estimation of dynamic models with a time series of repeated cross-sections", *Journal of Econometrics* 59(1): 99-124.

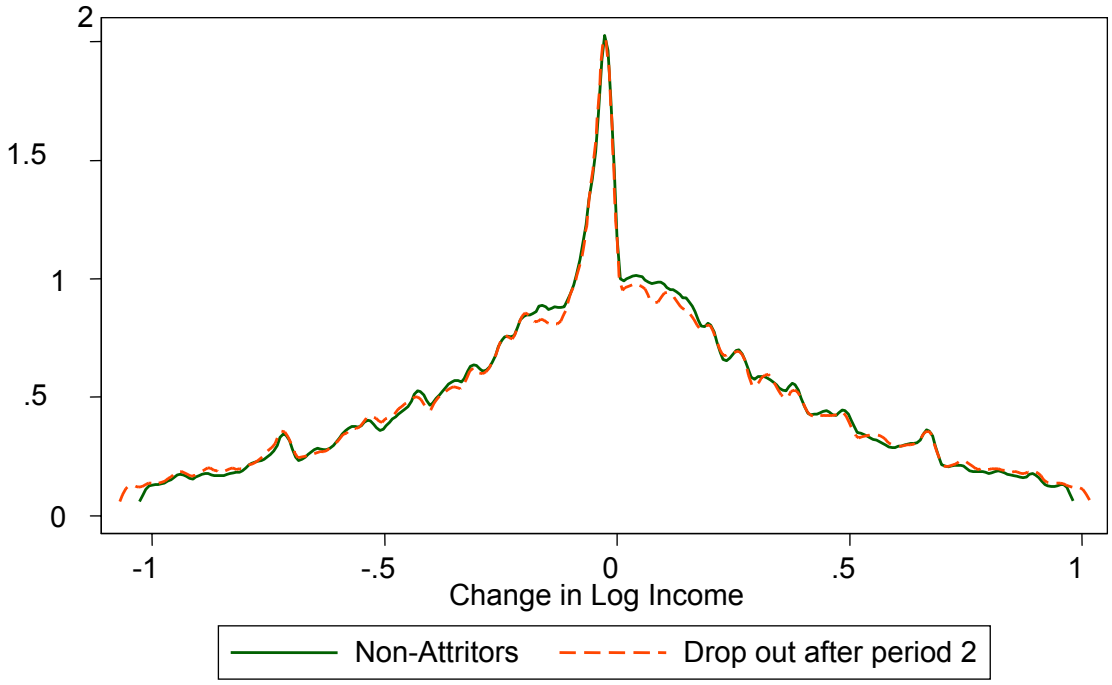
Piketty, Thomas (2000) "Theories of Persistent Inequality and Intergenerational Mobility", Chapter 8, pp. 429- 476 in A.B. Atkinson and F. Bourguignon (eds.) *Handbook of Income Distribution Volume 1*, North-Holland: Amsterdam

Strauss, John, Kathleen Beegle, Agus Dwiyanto, Yulia Herawati, Daan Pattinasarany, Elan Satriawan, Bondan Sikoki, Sukamdi and Firman Witoelar (2004) *Indonesian Living Standards before and after the financial crisis: Evidence from the Indonesia Family Life Survey*, RAND Corporation: Santa Monica.

Thomas, Duncan, Elizabeth Frankenberg and James P. Smith (2001) "Lost but not Forgotten: Attrition and Follow-up in the Indonesia Family Life Survey", Journal of Human Resources 36(3): 556-592.

Verbeek, Marno and Francis Vella (forthcoming) "Estimating dynamic models from repeated cross-sections", *Journal of Econometrics*.

Figure 1: Kernel Density of Change in Log Income between Periods 1 and 2 by Attrition Status



Kernel Densities plotted in STATA using the Epanechnikov kernel with a bandwidth of 0.01.

TABLE 1: COMPARISON OF ESTIMATES FOR QUARTERLY ENEU DATA

PANEL A: LOG SPECIFICATION
Dependent Variable: Log Real Household Income

	(1) OLS	(2) OLS	(3) IV	(4) A-IV	(5) Pseudo Panel	(6) Pseudo Panel
Quarterly Lag of Household Income	0.668	0.598	0.990	-0.062	0.988	0.832
T-statistic	955.85	791.23	213.59	-2.36	159.14	45.25
[95% confidence interval]	[.666,.669]	[.597,.600]	[.981,.999]	[-.114,-.011]	[.976,1.000]	[.796,.868]
P-value	0.00	0.00	0.00	0.02	0.00	0.00
Cohort Effects	No	Yes	No	---	No	Yes
Household-quarter observations:	1113172	1112464	165275	757561	---	---
Cohort-quarter observations:	---	---	---	---	842	842
Adjusted R squared:	0.4508	0.4731			0.9679	0.9703

PANEL B: LEVELS SPECIFICATION
Dependent Variable: Real Household Income

	(1) OLS	(2) OLS	(3) IV	(4) A-IV	(5) Pseudo Panel	(6) Pseudo Panel
Quarterly Lag of Household Income	0.376	0.316	0.999	0.012	0.973	0.738
T-statistic	447.64	366.61	88.83	4.97	103.64	32.86
[95% confidence interval]	[.375, .378]	[.315, .318]	[.977, 1.021]	[.007, .017]	[.954,.991]	[.694,.782]
P-value	0.00	0.00	0.00	0.00	0.00	0.00
Cohort Effects	No	Yes	No	---	No	Yes
Household-quarter observations:	1147860	1147127	169193	787094	---	---
Cohort-quarter observations:	---	---	---	---	842	842
Adjusted R squared:	0.1486	0.1862			0.9274	0.9362

Notes:
IV uses education of the household head and an asset index as instruments for lagged income; only for households observed in the third period.
A-IV denotes the Arellano (1989) instrumental variables estimator, which differences the data and uses $Y_{i,t-2}$ as an instrument for the lagged first difference.
All cohort-period observations are averages based on at least 100 household observations

TABLE 2: MOBILITY OVER DIFFERENT TIME INTERVALS
Pseudo-Panel Estimates from the ENEU

Dependent Variable: Log Real Household Income

	(1) Quarterly	(2) Yearly	(3) 2-Year	(4) 5-Year	(5) Quarterly	(6) Yearly	(7) 2-Year	(8) 5-Year
Lagged Log Household Income	0.988	0.963	0.936	0.950	0.855	0.536	0.051	-0.498
T-statistic	139.59	76.8	50.75	31.26	34.57	13.44	1.12	-13.63
[95% confidence interval]	[.974,1.002]	[.938,.987]	[.899,.972]	[.890,1.010]	[.807,.904]	[.458,.614]	[-.038,.140]	[-.570,-.427]
Cohort Effects	No	No	No	No	Yes	Yes	Yes	Yes
Cohort-quarter observations:	477	477	477	477	477	477	477	477
Adjusted R squared:	0.9762	0.9253	0.844	0.6722	0.977	0.9406	0.9174	0.9411

TABLE 3: WHO ATTRITS?

Probability of Attriting after first interview and Probability of dropping out anytime after second interview

	Only Present in First Wave	Attrition between Wave 2 and Wave 5			
	Attrit[a]	Dropout[b]	Dropout[b]	Dropout[b]	Dropout[b]
Independent Variables:	dF/dX	dF/dX	dF/dX	dF/dX	dF/dX
Initial Income of Household	8.05E-08	1.38E-06	8.59E-07	5.86E-07	7.06E-07
	(9.52)	(32.42)	(26.94)	(13.18)	(22.97)
Change in Income between periods 1&2 *	---	7.96E-07	0.003	2.39E-07	0.014
		(20.69)	(6.67)	(5.57)	(20.39)
Age of Household Head	-0.003	-0.020	-0.020	-0.020	-0.020
	(14.65)	(37.36)	(36.49)	(37.02)	(36.60)
Age Squared of Household Head	3.020E-05	2.135E-04	2.122E-04	2.118E-04	2.123E-04
	(11.05)	(29.37)	(28.73)	(29.14)	(28.75)
Household Size	-0.003	-0.014	-0.013	-0.013	-0.013
	(20.73)	(34.25)	(30.42)	(31.87)	(30.27)
Number of Children in Household	0.002	0.007	0.006	0.006	0.006
	(9.56)	(15.39)	(12.67)	(13.30)	(12.48)
Male	0.005	0.020	0.023	0.021	0.023
	(12.64)	(19.20)	(20.78)	(19.69)	(20.66)
Married	-0.009	-0.041	-0.043	-0.041	-0.042
	(24.55)	(44.50)	(44.42)	(44.15)	(44.08)
Education Dummies[d]	Yes	Yes	Yes	Yes	Yes
Number of Observations	1610094	1468273	1423132	1468273	1423132
Pseudo R squared	0.0128	0.0145	0.0141	0.0142	0.0144
Change in Probability due to 1 SD rise in initial income[c]	7.95E-04	1.37E-02	8.50E-03	5.76E-03	6.96E-03
Change in Probability due to 1 SD rise in change in income variable	---	8.02E-03	2.17E-03	2.25E-03	6.39E-03
*Change in Income Variable	---	Inc_2-Inc_1	Ln(Inc_2)-Ln(Inc_1)	Abs(Inc_2-Inc_1)	Abs[Ln(Inc_2)-Ln(Inc_1)]

Notes

[a] Attrit=1 if household only present for 1st interview; Attrit=0 if household present for longer than 1st interview

[b] Dropout=1 if household left anytime after 2nd interview; Dropout=0 if household present for entire panel (5 interviews)

Absolute value of Z-statistics in parentheses

[c] Evaluated at the means of all other variables

[d] Coefficients on education dummies are negative and statistically significant for all dummies in all specifications. Omitted group is no schooling.

TABLE 4: ARE ATTRITORS MORE MOBILE?
Quarterly Pseudo-Panel Estimates from the ENEU

Dependent Variable: Log of Real Household Income

	(1) Full Sample	(2) Non-Attritors	(3) Attritors	(4) Full Sample	(5) Non-Attritors	(6) Attritors	(7) Non-Attritors	(8) Attritors
Quarterly lag of Log Income	0.991	0.987	0.977	0.857	0.788	0.789	0.821	0.754
T-statistic	157.39	132.22	111.44	43.68	46.84	47	36.51	29.66
Confidence interval	[.979,1.003]	[.972,1.001]	[.960,.994]	[.819,.896]	[.755,.821]	[.756, .821]	[.777,.865]	[.704,.804]
Cohort Effects	No	No	No	Yes	Yes	Yes	Yes	Yes
Cohort effects restricted to be equal for attritors and non-attritors:	---	---	---	---	Yes	Yes	No	No
Cohort-quarter observations:	638	638	638	638	638	638	638	638
Adjusted R squared:	0.9749	0.9648	0.9512	0.9763	0.9617	0.9617	0.9673	0.956

Notes
Attritors are households that participated in all 5 quarters of the survey
Non-attritors are households that did not participate in all 5 quarters of survey
Columns (5)&(6) were run as one regression where only the slope coefficient was allowed to differ
Columns (7) & (8) were run separately as two regressions

TABLE 5: MOBILITY IN INCOME AND EXPENDITURE
Pseudo-panel estimates from the ENIGH Survey for two year lag

PANEL A: ABSOLUTE MOBILITY (NO COHORT FIXED EFFECTS)

	URBAN				RURAL			
	Log Primary Wage Income	Log Total Monetary Income	Log Full Income	Log Full Expenditure	Log Primary Wage Income	Log Total Monetary Income	Log Full Income	Log Full Expenditure
Two-year Lag	0.870	0.880	0.859	0.888	0.653	0.641	0.700	0.797
(T-statistic)	(14.69)	(14.45)	(13.20)	(15.08)	(5.07)	(4.91)	(5.52)	(7.06)
[confidence interval]	[.75,.99]	[.76,1.00]	[.73,.99]	[.77,1.01]	[.39,.91]	[.38,.91]	[.44,.96]	[.57,1.02]
Cohort-Period Observations:	61	61	61	61	40	40	40	40
Adjusted R^2	0.782	0.776	0.743	0.791	0.388	0.372	0.431	0.556

PANEL B: CONDITIONAL MOBILITY (COHORT FIXED EFFECTS)

	URBAN				RURAL			
	Log Primary Wage Income	Log Total Monetary Income	Log Full Income	Log Full Expenditure	Log Primary Wage Income	Log Total Monetary Income	Log Full Income	Log Full Expenditure
Two-year Lag	0.080	0.061	0.170	0.282	-0.078	-0.032	0.142	0.093
(T-statistic)	(0.52)	(0.37)	(0.98)	(1.73)	(0.44)	(0.17)	(0.74)	(0.47)
[confidence interval]	[-.23, .39]	[-.28,.40]	[-.18,.52]	[-.05,.61]	[-.43,.28]	[-.42,.36]	[-.25,.53]	[-.31,.50]
Cohort-Period Observations:	61	61	61	61	40	40	40	40
Adjusted R^2	0.838	0.827	0.780	0.824	0.611	0.540	0.530	0.659

Notes:
Absolute value of pseudo-panel t-statistic in parentheses.
Cohorts are defined by 5 year birth interval and three education groups.
Source: own calculations from the 1992-2002 ENIGH surveys.