Income Dynamics in Rural India:

Testing for Poverty Traps and Multiple Equilibria

Stefan Dercon

Ingo Outes[1]

June 2009 (revised)

Abstract

Do serious climatic shocks lead to processes of persistent poverty and poverty traps? We have access to a panel data set spread across 30 years, building on the ICRISAT data on six villages in the semi arid tropics to investigate this question. We identify the dynamic income process showing the transition dynamics in response to rainfall shocks, using a fixed effects dynamic model allowing for multiple equilibria. We show that there is serious persistence and evidence of multiple equilibria in the data: in the data period analysed, many households were initially in a precarious (unstable) equilibrium that could lead to a downward cycle into destitution, but also take-off if appropriate circumstances presented themselves. Many managed to escape in the period 1984-2004 towards higher and stable equilibria, leading to considerably better living conditions. For the median income houseold, the higher equilibrium is at roughly 155 US dollars per year per adult at 1976 prices, and the unstable equilibrium below which a downward spiral would emerge is about 55 US dollars. By investigating the fixed effects, we find that those with higher assets, especially in the form of initial levels of education in the family in the 1970s, higher land holdings and/or high physical capital were faced with a much lower level of income at which a downward spiral could have followed. Those with few assets in these different forms could experience the downward spiral at much higher levels of income: their livelihood was far more precarious. For them, climatic shocks, even at reasonable levels of incomes in preceding years could lead to destitution.

(Methodologically, we use Lokshin and Ravallion's estimation method but using rainfall as instruments rather than black box dynamic identification methods as in Arrellano-Bond estimators. The instruments are reasonably strong, and the link between climate shocks and destitution appear to be very strong. We find that rainfall-induced lower income does not only have a simple contemporaneous effect, as would be the case if rainfall caused the error in an income process that would be described as independently distributed errors. Instead, we find that rainfall-induced income levels have a persistent impact, possibly causing destitution.)

[1] Authors address: stefan.dercon@economics.ox.ac.uk and ingo.outes-leon@economics.ox.ac.uk. We are grateful to Joseph S. Shapiro for excellent research assistance. We also thank Reena Badiani for extensive and insightful help in preparing and processing the data.

1 Introduction

With about 2.5 billion people receiving less than two dollars per day in income (Chen and Ravallion, 2008), the issue of whether and why the destitute escape poverty constitutes a central question in economic research. Theories of poverty traps explain why living in poverty at some time causes a person to remain poor in the future, or why a country's poverty causes the country to remain in future poverty (Galor and Zeira, 1993). These theories imply stark conclusions: a positive income shock could prevent a person from living in poverty for the indefinite future, while a sufficiently grave negative shock to income could prevent a person from ever escaping poverty. Some such theories assume that a person requires a fixed and indivisible investment to purchase a good like education or credit (Banerjee and Newman, 1993); others assume increasing returns to income via nutrition or another means (Dasgupta and Ray, 1993); while still others show how leaving the poor without bargaining power can cause the poor not to save (Mookherjee and Ray, 2002). In this paper, we offer a test for the existence of poverty traps using long-term panel data from India.

Admittedly imperfect tests of these elegant models have offered little empirical support, however, leaving Dasgupta (1997) to describe that they reside 'awkwardly' in development thinking. A model of nutrition poverty traps has received empirical criticism from several studies (Bliss and Stern, 1982; Swamy, 1997; Rosenzweig, 1988), though Dasgupta (1997) argues that they use flawed tests. A theory of fixed costs to entering businesses has received similarly little support (McKenzie and Woodruff, 2003).

Several recent studies have proposed that a poverty trap could arise through a combination of mechanisms, or through some unstudied mechanism. These studies essentially examine whether a regression of some welfare measure (income, consumption, or assets) on its lag has a shape that could indicate the presence of a poverty trap, as described in figure 1, linking the welfare measure Y at t and t-1 in some non-linear way. A 45 degree line is drawn in to show equilibrium points, i.e. where Y at t is the same as at t-1. As is well-known, the shape shown offers multiple equilibria, and of those shown, a and c are stable (low and high) equilibria, and b is an unstable equilibrium, as once removed from b, a household would drift towards a or c according the dynamic relationship shown in the figure.

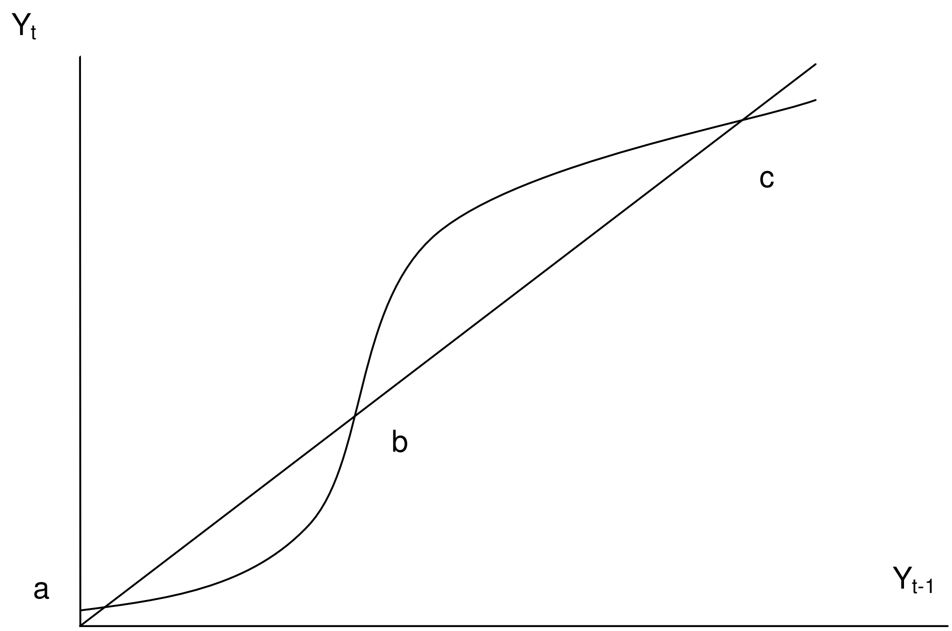


Figure 1: Poverty trap and multiple equilibria

Nonparametric kernel regressions of current on lagged assets using small samples from Kenya, Ethiopia, Madagascar, and South Africa show unstable equilibria over some low values of income that suggest the presence of a possible poverty trap. These studies ignore the endogeneity of lagged income in a dynamic panel models, however, and the potential bias in data obtained from many-year recall questions, limited generalizability of sample sizes under 200 individuals, and bias of bivariate kernel regressions at discontinuities (Fan, 1992) give their conclusions limited scope. In higher income areas, studies applying methods with corrections for various econometric challenges in estimating income dynamics to data from Eastern Europe, and Urban Mexico have found evidence for some stable low-level equilibria but no evidence of a poverty trap; evidence from China shows similar findings (Antman and McKenzie, 2007; Lokshin and Ravallion, 2004; Jalan and Ravallion, 2003).

The econometric challenges involved in testing for the presence of poverty traps are not trivial, and most create bias towards failing to reject the hypothesis that poverty does not entrap people. Hence one could reasonably conclude that the existing literature fails to establish whether poverty traps actually do not exist or whether available data and methods have inadequate power to detect them as there are econometric problems abound. Panel data with short duration--typically less than five years (Dercon and Shapiro, 2007) --may not capture the dynamics that ensure poverty's persistence. The nature of a dynamic panel model ensures that regression of income on its one-period lagged value will inflate the effect of lagged income on current income. Measurement error in income creates a mirage of income mobility, so a person whose true income remains constant over time may appear to enter then escape poverty. Similarly the stochastic nature of income might result in equally high levels of mobility (Barrett and Carter 2006).

Existing studies address some but not all of these concerns. Jalan and Ravallion (2003) and Lokshin and Ravallion (2003) use the Arellano-Bond GMM (Arellano and Bond, 1991) estimator to identify the association of a cubic polynomial of lagged income with current income. But if measurement error has serial correlation, as at least one U.S. comparison of survey-reported income with independent income reports suggests (Bound and Kruguer, 1991), then using distant lags of income as instruments for once-lagged income, as Arellano-Bond methods do, will overstate mobility. Antman and McKenzie (2007) for this and other reasons condemn the possibility of using panels for identifying nonlinear income dynamics, and propose instead the use of pseudo-panels to average out measurement error across individuals.

The present study shows how panel methods can address these econometric criticisms and consistently test for the presence of a poverty trap. We test whether a poverty trap characterizes the income dynamics of individuals in an unusually long 30-year panel from six villages in India's semi-arid tropics, based on a recent extension of the ICRISAT Village Level Studies. We follow (Lokshin and Ravallion, 2004) and others in estimating a dynamic equation where income is modelled as a cubic polynomial of lagged income that allows for unobserved income heterogeneity. However, unlike these studies, we use exogenous instruments to correct for endogeneity problems. We interact rainfall shocks with household characteristics to provide valid and informative instruments for a polynomial function of lagged income, obviating the need for Arellano-Bond methods and addressing the critical problem of measurement error in income.

Further, the econometric specification allows for household heterogeneity in income dynamics. We retrieve these individual effects and explore its correlates with starting period household characteristics, therefore uncovering those household assets that could have led to sustained increases in the trajectory of incomes.

What may look like an econometric solution to statistical problems, the method we use has clear conceptual meaning as well. Theoretical models of poverty traps typically imply that only a ‘shock’ can move people between equilibria, as shown in figure 1. In semi-arid India, from which the data are derived, the key is rainfall, and using rainfall as an instrument, we aim to make a direct causal link between rainfall as a cause of lower or higher income, affecting whether the household experiences a shock high enough to move between equilibria and the speed by which it moves to a new equilibrium.

Furthermore, the household fixed effects will allow households to have different underlying equilibria, reflecting for example different assets and human capital levels, offering a further interpretation on the meaning of the precariousness and potential in their livelihoods.

Our analysis shows that income generating dynamics in rural India follow a quadratic polynomial function with pronounced concavity. We find these results to be robust to sample trimming and changes in the period of analysis. Further, point estimates, obtained applying the ‘weak IV’-robust Fuller and LIML estimators, remain consistent with our original results.

Income simulations based on the estimated parameters suggest the presence of two equilibria: a stable high-income equilibrium and a low-level unstable saddle point. While households with sufficiently high fixed effect income follow the high-equilibrium dynamic path, almost half of our sample has too low an income steady state to overcome the dynamic point of divergence. Unpacking the household-specific fixed effects, we find that household assets in the early mid-1970s, especially higher levels of education, land holdings and physical assets, are positively associated with higher levels of steady state income.

Our analysis suggests, that over the past 30 years many households have managed to escape towards higher and stable equilibria leading to considerably better living conditions. However, those with few assets to start with could experience the downward spiral at much higher levels of income: their livelihood was far more precarious. For them, climatic shocks, even at reasonable levels of incomes in preceding years could lead to destitution.

The paper proceeds as follows. Section 2 outlines the econometric obstacles inherent in testing those models of poverty traps which only consider income dynamics. Section 3 describes the 30-year panel data set. Section 4 presents the main results, and section 5 concludes.

2 Econometric Methodology

An estimate of how lagged income affects current income must address four statistical problems: the endogeneity of lagged income in a dynamic panel model; measurement error in income; individual heterogeneity; and short panel duration. We discuss solutions for each, and we combine responses to these potential biases in the final estimator that we propose and implement.

2.1 Dynamic panels and measurement error

We estimate an AR(1) model where the income y_{it} of person i at time t depends on a linear function of a polynomial of person i 's lagged income,[2] and a composite error term with time-invariant and idiosyncratic components ρ_i and v_{it}:

$$y_{it} = \beta y_{i,t-1} + \rho_i + v_{it} \tag{1}$$

Since lagged income correlates positively with the composite error term $\rho_i + v_{it}$, estimating equation (1) by OLS generates inconsistent estimates of β. A first-differenced version of equation (1) eliminates the individual effect ρ_i:

$$\Delta y_{it} = \beta \Delta y_{i,t-1} + \Delta v_{it} \tag{2}$$

where $\Delta x_t = x_t - x_{t-1}$. However, the dynamic nature of the income generating process implies that OLS estimation of equation β in (2) will still produces inconsistent estimates, resulting from lagged errors ($v_{i,t-1}$) affecting both Δv_{it} and $\Delta y_{i,t-1}$ (Nickell, 1981).

Arellano-Bond IV General Method of Moments (GMM) estimators use lags of the endogeneous variable as instruments for $\Delta y_{i,t-1}$. If the error term Δv_{it} in equation (2) lacks second-order serial correlation, and if equation (1) is dynamically complete, then the second lag and any further lags of income can serve as valid instruments of $\Delta y_{i,t-1}$(see Arellano and Bond (1991) for the Arellano-Bond *dif* version of the estimator). Under these assumptions, the general method of moments provides

[2] Some studies describe such an estimate as a test of ``non-linear income dynamics'' (Antman and McKenzie, 2007; Jalan and Ravallion, 2003). While specifying lagged income as a higher-order polynomial allows current income to vary

consistent and effici;ent estimates of β, the parameter of interest (Anderson and Hsiao, 1982; Arrellano and Bond, 1991; Blundell et al. 2000; Bond, 2002).

Several existing studies use Arellano-Bond type of estimators with short panels to test whether a poverty trap characterizes income. One of these studies shows overidentification tests of the instruments' validity and a test that the residuals have second-order autocorrelation (Jalan and Ravallion, 2003); the other studies do not mention these tests Lokshin and Ravallion, 2004; Antman and McKenzie, 2007], and none of the three studies evaluates the weakness of the instruments, an important problem when a dynamic panel has a near-unit root (Stock et al., 2002, Blundell et al, 2000), and a problem which large samples do not eliminate (Bound et al., 1995).

Measurement error also creates a more substantial problem in these papers, since some U.S. data suggest that measurement error in an individual's income has positive autocorrelation across waves of a panel (Bound and Krueger, 1991). Antman and McKenzie (2005) show that in the presence of such measurement error, GMM estimators provide inconsistent estimates of the parameters in equation (2), and severity of the problem persists upon specifying lagged income as a higher-order polynomial.

When applying Arellano-Bond methods, measurement error, the stochastic nature of income, as well as poor instrument validity of higher-lags of income will all conspire in creating the mirage of an income generating process with high levels of mobility. In this paper, instead, we propose the use of exogenous instruments in the form of rainfall shocks.

Rainfall provides a useful instrument for income in poor agricultural areas.[3] Since the economies of agricultural villages heavily depend on weather, flood and drought sharply affect the income of most households in the village: agricultural households have lower yield in seasons of extreme weather, households that earn income from agricultural labour find less work in times of extreme weather, and most individuals in these communities depend on good weather for strong income.

Although our rainfall data is only available at the village level, we expect precipitation shocks to have heterogenous effect across households. On the one hand, positive and negative rainfall shocks will have different effects across households with different levels of reliance on agricultural income. At

---

nonlinearly with lagged income, the regression function itself is linear in the higher-order terms of lagged income.

[3] See, for example, Paxson (1992), Miguel (2005), and the review in Rosenzweig and Wolpin (2000).

the extreme, land cultivating households would benefit more from good rain than landless households. On the other hand, in economies of household production with imperfect labour markets such as rural India, labour-scarce households will not be able to make the most of a particularly good rainy season.

We use three different exogenous instruments. First, we use as rainfall shock deviations in percentage of a year's precipitation relative to its historic mean. For this we use monthly village-level rainfall measures from several Mandal-level collection stations. Second, we interact this rainfall shock variable with the land area (in hectares) operated by the household in that year. Size of land usage makes intuitive sense as an instrument, since households with larger plots will receive more benefit from years with good rainfall and more harm from years with drought. We use as a third instrument the interaction between rainfall shocks and the number of children aged 0-8, as a measure of the labour-scarcity of the household. Households with more inactive members too young to contribute and requiring care will be less flexible to insure their income against rainfall shocks.

For each rainfall instrument $Z_{i,t-1}$, we require two conditions to be met:

$$cov(Z_{i,t-1}, \Delta y^*_{i,t-1}) \neq 0 \qquad (3)$$

$$cov(Z_{i,t-1}, \Delta v_{i,t}) = 0 \qquad (4)$$

where $y^*_{i,t-1}$ denotes true, unobserved income. Condition (3) requires that the instrument strongly correlates with true lagged income – the 'strong' IV condition –, while condition (4) requires the instrument to be orthogonal to measurement error in lagged income or with other components of the structural equation error term – the so-called 'validity' condition. One could interpret our critique of Arellano-Bond estimates of equation (2) as violations of the validity condition.

While we expect our rainfall instruments to meet this condition, violations of assumption (3) might pose a challenge to our results. Problems arising from 'weak' IVs are particularly severe when one considers the finite-sample properties of IV estimators. Point estimates are rendered biased and inconsistent, while standard errors are invalid (see Staiger and Stock (1997), Hahn and Hausman (2005) and Murray (2006a) among others). As a robustness check on the IV GMM estimates, we apply 'weak' IV-robust estimators. Fuller k-class estimators and limited information maximum likelihood (LIML) estimators are understood to perform better under 'weak' IVs.[4] Furthermore, in a world of weak IVs, the potential IV bias can be reduced when using a parsimonious set of instruments (See Stock and Yogo (2005)). We re-estimate our model with a reduced set of instruments. This will not only provide a further robustness check on our main results, but it can provide valuable information on the direction of the remaining bias in our point estimates.

2.2 Individual heterogeneity

It is possible that individuals have multiple equilibria for income, or that poverty creates a trap for some but not all individuals. Fixed individual factors -- education, geographic location, and others -- may affect the trajectory of an individual's income. Since these fixed factors may correlate with income and hence bias regression estimates, equation (2) uses first-differencing to eliminate these fixed effects.

But the effects themselves have economic interest, and recovering these parameters allows us to observe the correlation between observable individual fixed characteristics and the part of an individual's income trajectory which does not depend on short-term income dynamics. It will give us insight in any factors that may cause household incomes to increase by a certain amount each year, and the individual effects contain these effects. Furthermore, as will be shown further below, the fixed effects will affect the location of the equilibrium, and we can give then meaning to the type of households that have higher equilibria compared to others. Jalan and Ravallion (2005) and Antman and McKenzie (2005) estimate models with household fixed effects, but fail to examine further its correlates, even though they both highlight their relevance and how they shift a person's trajectory.

Since the idiosyncratic errors have mean zero across the population, we estimate the individual effect by the deviation of an individual's mean outcome from the predicted mean (Antman and McKenzie, 2007):

$$\hat{\alpha}_i = \overline{Y}_i - \hat{\beta}_1 \overline{Y}_{i,t-1} - \hat{\beta}_2 \overline{Y}^2_{i,t-1} - \hat{\beta}_3 \overline{Y}^3_{i,t-1}$$

where we average the dependent and independent variables across the years in which they would

[4] Studies have shown that the Fuller k-class of methods dominates over other 'weak' IV-robust estimators. However, LIML methods are also often used for its nesting properties: when the model is exactly identified GMM and LIML are identical, and Fuller estimates are mean-square-error corrected versions of LIML. See Hahn, Hausman, and Kuersteiner

appear if we had not first-differenced the model.

To estimate the correlates of these fixed effects, we regress them on a vector Z_i of fixed individual characteristics:

$$\alpha_i = \phi_0 + Z_i\phi_1 + \varepsilon_i$$

The parameters ϕ_1 show the correlation of individual characteristics with the fixed effects. A positive association $\phi_j > 0$ for some element j of the vector Z_i implies that ϕ_j gives an individual continuously increasing income regardless of shocks. Since we observe income and several fixed characteristics each wave only at the household level, all regression estimates in the paper use standard errors robust to heteroskedasticity and serial correlation within household-years.

2.3 Panel Duration and Stability of Income Dynamics

The dataset we use has the advantage of an unusually long duration: 30 years, a length paralleled by only a small handful of existing datasets (Dercon and Shapiro, 2007). Unfortunately the panel has a gap of about fifteen years: households were surveyed yearly between 1975 and 1983, which we name the VLS1 panel, and again between 2001 and 2005 – the VLS2 panel (see the following section).[5] In the effort to examine the long-run factors that influence poverty and welfare, such long-term panel duration provides critical information on income dynamics. But given our focus on income dynamics, ignoring this gap in the middle and treating 1983 as if it preceded 2001 will yield problematic estimates for later years.

To address the 1984-2000 gap, we use one-year lags of variables for all years. Taking the first difference of income model with one-period lags as a repressor, drops two years in each of the two period panels. Hence, in our model specification we use the first difference of current income from nine waves of the panel (1977, 1978, 1979, 1980, 1981, 1982, 1983, 2003, 2004), while we use the first difference of the independent variables (lagged income) from a different set of nine waves (1976, 1977, 1978, 1979, 1980, 1981, 1982, 2002, 2003). Since we use lagged rainfall as an instrument rather than the many lags of income used by Arellano-Bond estimators, the 1984-2001

(2003), Anderson, Kunitomo, and Matsushita (2005) and the review article Murray (2006b).

[5] Income data in year 1984 included only a small subset of individuals. A 1992 round of income data included few individuals and had different methodology than other years, while 2005 data are still being processed.

gap creates no other obstacles in estimating the dynamic panel model.

The 15-year gap between the two annual survey panels raises a further concern. Our model specification assumes that a single set of polynomial parameter values underlie the income generating process. Similarly, the model assumes that the household-specific fixed effects are stable over time. However, should the income dynamics in rural India have changed sufficiently during the 15-year gap between VLS1 and VLS2, our model specification might be grossly mis-specified resulting in uninformative parameter estimates. We address this issue by testing the parameter stability of the income polynomial function across the two panel periods.

Requiring a very long panel comes at at least one cost: attrition. Over a 30-year period, a considerable number of households were lost, partly due to the well-documented rules of tracking the ICRISAT panel (Foster and Rozenzweig, 2001), which in principle did not track anyone leaving the household. In the new rounds, split-offs were largely included, but migration is not irrelevant, and the main cause of attrition (Badiani et al., 2007). The sample is therefore a sample of households that is living in the respective villages in 2001-05 that are directly related to households already in these communities in the 1970s. This obviously affects external validity, but as long as coefficients are interpreted exactly as relevant for this sample, there is not a problem, and obviously, given the relative low mobility in rural India (Munshi and Rosenzweig, 2009), this is not an irrelevant sub-population. Furthermore, the analysis in this paper is effectively to uncover a 'technological' relationship in the income process faced by a particular population: do poverty traps exist for them.

3 Data: the 30-year ICRISAT Panel

The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) near Hyderabad, India, collected annual surveys between 1975 and 1984 (VLS1), then for the same households in the period 2001-2005 (VLS2). The core data included 60 households each from six villages (240 in total) in India's semi-arid topics: the villages of Aurepalle and Dokur in the Mahbubnagar District of the Indian state of Andhra Pradesh; the villages Shirapur and Kalman in the Sholapur District of the state of Maharashtra, and the villages Kanzara and Kinkheda in the Akola District of Maharashtra. Villagers generally work in dryland farming, with limited irrigation (Badiani et al. 2008).

For the early data collection, interviewers lived in the villages and interviewed households every 3-4 weeks to obtain income information. The more recent data (VLS2) use one interview per year for 2001-2003 and two per year for 2004. Detailed checks on comparability between these years and with the VLS1 is reported in Badiani et al. (2008). A tracking survey allowed follow-up of individuals interviewed in the 1975-84 rounds. Additionally, the 2001-2005 study re-surveyed new households to compensate for the reduced sample sizes due to attrition. Walker and Ryan (1990) provide detailed description of the early survey rounds and research stemming from them, while Badiani et al. (2008) provide an appendix with further detail on the recent data collection. A key point to notice is that this paper provides with an assessment of the comparability of different indicators, as the frequency of data collection is different in the VLS1 and VLS2. Badiani et al. (2008) show that trends in those variables that were collected with the same frequency in both surveys and those that were not showed remarkable similarity suggesting that comparability may not be negatively affected.

Table (1) provides descriptive statistics of the general structure of the data. As one would expect of a 30-year long panel, household attrition is substantial. One reason was that especially in the first few years, a considerable number of households (more than 10 percent) dropped out, and were replaced. As for many of these replacements we have a long series, we have kept them. Migration and death meant that a considerable number dropped from the sample over time; in later years, split-offs still living in the village were included in the sample.

Interested in modelling income dynamics over a long period, we centre our attention on household for which we have sufficient information in both VLS1 and VLS2 panels. In terms of our models specification, this requires households to be measured at least for two continues waves in both annual panels. Our analysis therefore does not include original VLS1 households that were not resurveyed in the later years. Also, the samples were boosted in 2001, but these were not included either. However, we do include split-off households by constructing an income series for the full length of the panel based on matching their VLS2 income series to the VLS1 series of the households they originated from. The result is 260 households, linked to 151 'original' households that are used in this paper.

This is hardly a 'representative' sample of either the households living in the villages at baseline or in 2005, even though it started as a random sample for each village in 1975. However, given that all those initially included in the sample in the 1970s are included provided that they still were in the village in 2001, it will give us inference of whether for this particular type of household. It may be that some are trapped because they could not attrit via migration or that some that disappeared due to mortality did so because they were trapped in poverty. Our study cannot address this.

Finally, before moving to the discussion of our results, we note that throughout the paper standard errors reported correct for heteroskedasticity and autocorrelation in the error term for specific clusters. All regressions reported apply clustering at the level of the household. However, considering that some of our households are split-offs from an original VLS1 households, we have also carried out estimates where we correct for autocorrelation across sibling and parent households. We find results to be robust to this modification (results available upon request).

4 Results

4.1 Income Trends

Table [2] reports some descriptive statistics of our household income measure (household income per adult equivalent per year in rupees in 1976 prices). The early panel, the VLS1 sample, shows a slight upward trend, partly due to particularly low income in the first two years of the survey. Mean income in the later panel, VLS2, is substantially higher than in VLS1, more notably due to the 70 percent increase between years 1983 and 2001, a finding that coheres with the regression results of Badiani et al [2008] using the same data. Graph [1] provides an even more striking representation of the income increases benefiting the household in the ICRISAT. The graph plots kernel densities of income for individual years 1975, 1983, 2001 and 2004. We observe that not only has mean income increased over this period, but the spread of the income distribution has also grown dramatically.

Before moving to the discussion of the results of our parametric analysis, it is instructive to plot the raw income data. The econometric model assumes that the income generating process follows a polynomial function. Graph [2] uses locally weighted (Lowess) methods to obtain a non-parametric estimate of income lagged ($y_{i,t-1}$) on current income ($y_{i,t}$). The income patterns implied by the graph suggest substantial convergence and few non-linearities in the income generating process. In the next section, we assess whether this remains an accurate description of the true income generating process once issues of measurement error, income stochastic patterns and unobserved heterogeneity have been addressed.

4.2 Regression results

As discussed in section 3 above, our econometric methodology estimates the parameter of a polynomial function of lagged income in first-differenced form, equation (2), using exogenous instruments. Tables [3] and [4] report IV GMM first and second-stage estimates. We fit three alternative polynomial functions – linear, quadratic or cubic –, and for each of these model we estimate one specification with year dummies and another without.

Columns (1) and (2) in Table [3] show that our instruments are a good predictor of income; the joint significance of our instruments is well above 10, and the implied Cragg-Donald suggests the absence of weak instruments. When no year dummies are included, we find rainfall deviations significantly increase household income. Similarly, the interaction effects between rainfall shocks and household characteristics are also significant and take signs as expected. We find that households operating larger plots and households with fewer kids among their members appear to benefit (lose) the most when rainfall is abundant (poor).

While estimates for interaction instruments remain largely unchanged when year dummies are included, rainfall shocks lose their significance. With year dummies, the income effect of rainfall shocks is exclusively identified by the variation in village precipitation. Given that the pattern of rainfall in the Indian sub-continent is mostly determined by the timing and profusion of the monsoon season, we would expect rainfall shocks in a given year to be highly correlated across villages. It is therefore not surprising to find rainfall shocks not to vary sufficiently across villages to be identified beyond the annual effect. As other factors determining incomes may also be common across villages, a specification with time dummies would seem more parsimonious and therefore we focus in the rest of the paper of these results.

Further, evidence in columns (3) to (6) in Table [3] shows that the interaction instruments are not only good predictors of the first moment, but also of the second and third moments of income, although less strongly so than levels of income. Estimated coefficients both take signs that make economic sense and are significant at standard levels of confidence.

In summary, we find that our set of instruments is strongly correlated with income. In fact, when we test for their joint-significance we obtain a high F-Statistic for the first moment of income (11.47) and non-negligible values for the higher moments, 5.48 and 3.55 for the quadratic and cubic terms respectively. Even though encouraging, we cannot ignore inference problems arising from 'weak' instruments. We return to this issue in the next section.

In Table [4] we report our main results. As in Table [3], we report second-stage estimates for three different polynomial specifications – linear, quadratic and cubic – and two variations of each model, with and without year dummies. As mentioned before, here we focus on the results with year dummies.

As expected, when fitting a linear model we find a positive correlation between lagged income and current income (see column 2). The coefficient is statistically significant but modest in magnitude suggesting a relatively high degree of income mobility, or high speed of convergence.

Estimates for the quadratic and cubic model specifications provide some striking results. First, we find that the income generating process might not be linear in nature. When we include the second-moment of lagged income we find it to be significant at the 5% level – see column (4). It loses its significance when we allow for a third-moment in the income process – possibly due to multicollinearity.[6]

Secondly, the point estimates for the first-moment of lagged income in columns (4) and (6) are significant and very large in magnitude, suggesting convergence might be relatively slow or even unachievable. Indeed, point estimates when allowing for higher-order polynomials show that the income generating process might follow a concave pattern, opening the possibility for the existence of multiple equilibria. In the Appendix, we discuss a number of checks for robustness of these results, and generally, these findings are confirmed.

4.3 Multiple Equilibria and Household Heterogeneity

Showing that the current income follows a polynomial of lagged income with a concave pattern does not constitute proof of the existence of multiple equilibria. For that we need to show that derivative of the resulting polynomial is larger than unity when $y_{i,t}^{*} = g(y_{i,t-1}^{*})$ – where the g-function

[6] Indeed, the correlation coefficient between the second and third moment of lagged income in these regressions is very high, reaching 0.96 when estimating the model in column (6)).

represents the dynamic polynomial function – and that this condition is met within the range of values of the income distribution. Furthermore, in our model specification, household heterogeneity will shift the dynamic patterns and therefore any potential multiple equilibria will be household specific (see also Antman and McKenzie (2005)).

The easiest way to illustrate this is graphically. Graph [3] plots income simulations based on the parameter estimates of the polynomial from the quadratic model specification reported in Table [4].[7]

The results presented in Graph [3] are truly striking. First, when considering the dynamic path for the median household, we find two equilibria in the range of reasonable values of income: namely a high stable equilibrium, and a low unstable saddle point. The high stable equilibrium is approximately 1400 Rupees per adult per year in 1976 prices, approximately 155 US dollars at the exchange rates at that time. The lower unstable equilibrium is about Rp 500 or 55 US dollars. The derived dynamic path implies that households with fixed effect values close to the median will face a divergence point or threshold whereby the dynamic paths separate. Namely, household lying above the saddle point will converge over time towards the high equilibrium while households below the saddle point will inevitably suffer further losses in their future income. Lying below the threshold or being pushed over it by shocks such as rainfall would appear to put households on a path towards 'perdition'.

A second aspect to note from Graph [3] is that not all households are exposed to the risk of destitution. Household with sufficiently high individual fixed effects, as represented by the top decile in Graph [3], have a single dynamic equilibrium. For this type of households, we expect their income levels to converge – although the rate of convergence suggested by the concave polynomial would appear to be very slow, even for relatively high values of lagged income. Furthermore, the existence of a single equilibrium for this group of households cannot be understated. Even when faced with large shocks, these households would appear to face little risk of being put on a path of structural divergence. It is as if their livelihood faces no vulnerability: even if they occasionally have low incomes, they won't get stuck there permanently.

Thirdly, among households with low fixed effects, we find that the dynamic path also appears to

[7] We report in the Appendix the full set of simulations based on all six of the specifications estimated in Table [4]. After recovering the household fixed effects implied by the model estimates, we plot the dynamic simulations for different percentiles of the fixed effects distribution. Specifically Graph [3] reports the simulated dynamic path for the 10th, 50th and 90% percentile of the individual household fixed effects.

experience multiple equilibria. However, for this group vulnerability does have a qualitatively different meaning than for other households. For low levels of fixed effect income, the stable equilibrium and the divergence threshold are close to each other. In other words, for households that have already reached their steady state equilibrium, a relatively small shock could push them over the divergence threshold. Furthermore, it should be noted that the lower the fixed effects, the higher lies the saddle point. In other words, households with low fixed effects could be set on a structural divergence path even though they might have values of current income higher than households with higher fixed effects that are on a convergence path towards their stable equilibrium.

Two further points are worth noting. Analysing the median household dynamic path we see that the threshold for this group lays on a relatively low value of income. It is therefore plausible that most households might have current income above the threshold, resulting in a relatively low risk of structural destitution. If this applies to the median, it is possible that the risk of being set on a divergence path is only real for a relatively small sub-sample of households. However, Graph [3] provides some evidence against this possibility. Comparing the median versus the bottom decile households' curves, we see that there is only a small vertical distance between the two. This means that almost half of our sample of households face a dynamic threshold that lays between 200 and 750 Rupees – values that capture approximately [35%] of the lagged income distribution.

Additionally, although clear from the graph, it is instructive to recognise the fact that households on a convergence path will ultimatively converge to their own steady state equilibrium. The actual equilibrium income they will reach will therefore depend on their own household fixed effects.

4.4 Unpacking Household Fixed Effects

The importance of the individual household fixed effects can hardly be overstated. Not only do they determine the steady state income households will eventually reach, but households with fixed effects approximately below the median are faced with the real possibility of suffering a shock that might put them on a dynamic path towards destitution. It is with this in mind that we now move towards understanding what lays behind these fixed effects.

To retrieve some idea of the correlates of income fixed effects, we regress them on a set of starting period household characteristics as measured at the beginning of our panel in 1975. Columns (1) to

(6) in Table [5] report the correlates of fixed effect income as computed from models (1) to (6) in Table [4]. While these six columns include only time-invariant household characteristics, columns (7) to (12) add two additional time-varying household assets: land area owned and value of household assets.

Results shown in Table [5] are robust across our six different model specifications. For our preferred model, the quadratic polynomial with year dummies, we find that beyond village dummies, education of the household head appears to be significantly correlated with fixed effect income. When we add land ownership and value of assets to education, we find all three to be strongly correlated with the fixed effects. Despite their geographic proximity, these villages have substantial heterogeneity in soil and other characteristics (Walker and Ryan, 1990). Correspondingly, individuals in different villages have different income trajectories: compared to village 1, the reference, villages 3, 4 and 6 have substantially lower fixed effects.

We interpret these results as suggestive that while changes in India over the past 30 years have increasingly created opportunities for substantial welfare improvements, not all households have been in the position to benefit. Human capital and physical assets appear crucial in ensuring that households are well enough equipped to take up the opportunities.

5 Conclusions

A variety of theories suggest why a person who becomes poor at any time will remain poor indefinitely. Most such theories focus on a technology with increasing returns to scale which arises from a particular social mechanism--nutrition, education, fixed costs to entering a business, or another. The ideas of poverty traps that arise from these theories constitute a central theory of development economics at both the micro and macro levels. But these theories have received extremely little empirical support, possibly due to econometric pitfalls in the methods underlying the relevant empirical studies, as Dasgupta (1997) argues occurs for tests of the nutrition-efficiency wage theory, or possibly because no poverty trap in fact exists.

The large number of people in extreme penury constitutes only one reason underpinning the importance of understanding whether and why the destitute escape poverty. The presence of poverty traps would also implies a startling policy conclusion: a small transfer to a poor individual or household could change that person from low- to high-level equilibrium and permanently remove a person from poverty.

Since most existing theories of poverty traps assume some form of fixed investment cost, or increasing returns to assets or income, we examine whether income dynamics give evidence of increasing returns. A variety of econometric problems arise in this analysis: lagged income is inherently endogenous in a dynamic panel model; measurement error in income will cause OLS or GMM estimates to understate income's persistence; individual heterogeneity may disguise the fact that some individuals face a poverty trap even though the average individual does not; and short panel duration may give inadequate time to observe sufficient movement in income.

The bivariate kernel regressions or Arellano-Bond methods that existing papers use address some but not all of these pitfall. We apply IV GMM methods in a dynamic income equation that addresses issues of measurement error and endogeneity of lagged income, while allowing household heterogeneity in income steady state. Unlike similar studies (Jalan and Ravallion (2003) and Lokshin and Ravallion (2004)), we use exogenous instruments in exploiting deviations in annual precipitation to explain future income. Indeed, first-stage estimates reveal rainfall deviations to be a strong predictor of year-on-year changes. In particular, when we interact rainfall with household land operated and composition variables to obtain valid and relatively strong instruments for a polynomial function of contemporaneous income.

Our analysis shows that income generating dynamics in rural India follow a quadratic polynomial function with pronounced concavity. We find these results to be robust to sample trimming and changes in the period of analysis. Further, point estimates, obtained applying the 'weak IV'-robust Fuller estimator, remain consistent with our original results.

Income simulations based on the estimated parameters suggest the presence of two equilibria: a stable high-income equilibrium and a low-level unstable saddle point. While households with sufficiently high fixed effect income follow the high-equilibrium dynamic path, almost half of our sample has too low an income steady state to overcome the dynamic point of divergence. Analysis of income fixed effects shows that schooling and other assets at the beginning of the sample period are linked to high levels of steady state income.

We interpret our results as suggesting that changes in India over the past 30 years increasingly provide opportunities for substantial welfare improvements, but not all households are well place to benefit. Education appears crucial in ensuring these opportunities are being taken. Those with higher assets have an income process with a much lower low-level unstable equilbrium than those with fewer assets: the latter's lives are far more precarious and even at higher income levels they risk sliding down dramatically. For some with high assets, this low unstable equilibrium would correspond to large negative current income positions. While in an agricultural setting occassional negative incomes are possible (and indeed observed in the data), it suggests that only a rather high and almost improbable income draws they would face such outcomes.

Appendix

A.1 Estimation with Weak Instruments

As discussed in section 2, consistent IV GMM estimates require for two instrumental variable conditions to hold, the 'validity' condition and the 'strong IV' condition. The validity condition states that the set of instruments should not be correlated with any unobserved determinant of income. The exogenous nature of the rainfall shocks and its potential heterogeneous effects, suggest that our set of instruments is unlikely to be invalid. Indeed, Hansen J Overidentification statistic reported in Table [4] indicates that we cannot reject the null hypothesis that all excluded instruments are exogenous.

However, Cragg-Donald F-statistics related to Table [4] suggest that our set of instruments might not be sufficiently strong. Although our excluded instruments are good predictors of individual moments of lagged income, the Cragg-Donald statistics test the null that the set of excluded instruments is jointly sufficiently strongly correlated with the set of endogenous variables. We find that in spite of first-stage F-Statistics of 11.47, 5.48 and 3.55 for the three-moments of changes in lagged income, the Kleibergen-Paap rank corrected Cragg-Donald statistics amounts to values of 2.48 and 0.49 for the quadratic and cubic models with year dummies. When compared with Stock-Yogo critical values, these Cragg-Donald statistics suggest the presence of 'weak' instruments resulting in IV GMM estimates containing absolute biases approximately exceeding 30%.

Such magnitude of bias casts doubts on the reliability of the results presented earlier. In this section, we follow two alternative approaches designed to provide further evidence of the robustness of our earlier results. First, we apply alternative estimators that are more robust to 'weak' instruments. We use LIML and Fuller estimators to alternative point estimates for our main TSLS results. Secondly, in a world of weak IVs, the size of the resulting bias is understood to increase with the number of instruments. As a robustness checks we re-estimate our quadratic model, using only the interaction effects of the rainfall shocks. While removing the rainfall shock itself from the instrument set might reduce the bias, this comes at a little cost, as rainfall shock itself is not significant when including year dummies alongside.

Table [A1] in the Appendix reproduces Table [4] for the alternative LIML and Fuller estimators. Comparing these results with the IV GMM estimates, we find that, although significance is lost for some coefficients, LIML and Fuller estimators provide remarkably similar point estimates. These alternative estimators appear to suggest an income generating process that follows a quadratic polynomial with concave trajectories. Although still suffering from substantial weak IV bias, we draw some comfort from the fact that these alternative estimators provide point estimates consistent with IV GMM estimates.

Additionally, Table [A2] reports results for our ‘parsimonious IV’ estimates of the quadratic model with year dummies. Using only the rainfall shocks interacted with land area and with the number of kids in the household, we improve the strength of our instrument. Indeed first-stage F-Stats increase to 15.08 and 6.98 for first and second moments respectively, and the Cragg-Donald statistic reaches a value of 3.35. Although the latter is not sufficiently high for weak IVs to be ruled out, the revised estimates would be expected to contain a smaller bias.

Results reported in Table [A2] are broadly consistent with our earlier results. More interestingly from our perspective is to see that relative to LIML and Fuller estimates with three excluded instruments, points estimates increase in magnitude for both the first and second moment. We interpret this change as an indication that the ‘weak’ bias included in our TSLS estimates might be biasing downward the estimated parameters.

The weak IV bias also has consequences for our earlier results regarding simulation graphs. Graphs [A2] in the Appendix, reproduces the simulation dynamics for the regressions using the ‘parsimonious IV’ set reported in Table [A2]. While GMM estimates remain largely unchanged, we find that the lower magnitude in Fuller point estimates implies that the less pronounced concavity in the dynamic function leads to a single equilibrium model.

While this has the consequence that no divergence thresholds would apply for any household in our sample, regardless of their fixed effect size, concavity leads to a very slow rate of convergence. In other words, should we take our Fuller estimates at face value they would suggest that temporary unexpected shocks would have long and persistence effects on future income, even if it is not a 'trap'. Furthermore, the speed of recovery would appear would be slowest for households with lowest fixed effects.

A.2 Parameter Stability and Panel Structure

Our model specification assumes that a single set of polynomial parameter values underlie the income generating process. However, considering the 15-year gap between VLS1 and VLS2, it is plausible that income dynamics have dramatically changed during that time. We address this issue by testing the parameter stability of the income polynomial function across the two panel periods. Table [A3] reports results from estimating the quadratic model specification for the VLS1 and VLS2 panel samples separately. Although significance is lost in the short panel, point estimates are remarkably similar to the estimates for VLS1 only. Indeed a test of parameter equality between the two periods indicates that point estimates are not significantly different from each other. Similarly, dynamic simulations based on these coefficients suggest that income dynamics between the two periods are very similar.

References

Anderson, T.W. and Cheng Hsiao. Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*, 18(1):47-82, 1982.

Antman, Francesca and David McKenzie. Poverty traps and Nonlinear Income Dynamics with Measurement Error and Individual Heterogeneity. *Journal of Development Studies* 43 (6) 1057-1083, 2007.

Arellano, Manuel and Stephen Bond. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 48(2):277-297, 1991.

Badiani, Reena, Stefan Dercon, Pramila Krishnan and K.P.C. Rao. Changes in Living Standards in Villages in India 1975-2004: Revisiting the ICRISAT village level studies. CPRC Working Paper Series, no. 85, 2008.

Banerjee, Abhijit and Andrew Newman. Occupational Choice and the Process of Development. *Journal of Political Economy*, 101(2):274-298, 1993.

Bliss, Christopher J. and Nicholas H. Stern. *Palanpur: The economy of an Indian village*. Oxford University Press, Oxford, 1982.

Blundell, Richard and Stephen Bond and Frank Windmeijer. Estimation in dynamic panel data models: improving on the performance of the standard GMM estimators. 2000.

Bond, Stephen R. Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2):141-162, 2002.

Bound, John and David A. Jaeger and Regina M. Baker. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430):443-450, 1995.

Bound, John and Alan B. Krueger. The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?. *Journal of Labor Economics*, 9(1):1-24, 1991.

Chen, S. and Martin Ravallion. "The Developing World Is Poorer Than We Thought, But No Less Successful in the Fight against Poverty", Policy Research Working Paper 4703, The World Bank. 2008.

Dasgupta, Partha and Debraj Ray. Inequality as a Determinant of Malnutrition and Unemployment: Theory. *The Economic Journal*, 96(384), 1986.

Dasgupta, Partha, "An Enquiry into Well-Being and Destitution", Oxford University Press, 1993.

Dercon, Stefan and Joseph Shapiro. "Moving On, Staying Behind, Getting Lost: Lessons on poverty mobility from longitudinal data", chapter in Deepa Narayan and Patti Petesch, *Moving out of Poverty*, World Bank 2007.

Fan, Jianqing. Design-adaptive Nonparametric Regression. *Journal of the American Statistical Association*, 87(420), 1992.

Fitzgerald, John and Peter Gottschalk and Robert Moffitt. An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources*, 33(2):251-299, 1998.

Galor, Oded and Joseph Zeira. Income Distribution and Macroeconomics. *Review of Economic Studies*, 60(1):35-52, 1993.

Heckman, James J. Formulation and Estimation of Dynamic Models Using Panel Data. *Econometrica*, 47(1):153-162, 1979.

Jacoby, Hanan and Emmanuel Skoufias. Estimating the Return to Schooling: Progress on Some Persisitent Econometric Problems. *Review of Economic Studies*, 64, 1997.

Jalan, Jyotsna and Martin Ravallion. *Insurance Against Poverty*, chapter Household Income Dynamics in Rural China. Oxford University Press, 2003.

Lokshin, Michael and Martin Ravallion. Household Income Dynamics in Two Transition Economies. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.

McKenzie, David and C Woodruff. Do entry costs provide an empirical basis for poverty traps?.

BREAD Working Paper No. 020, 2003.

Mookherjee, Dilip and Debraj Ray. Contractual Structure and Wealth Accumulation. *American Economic Review*, 92(4):818-849, 2002.

Nickell, Stephen. Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417-1426, 1981.

Rosenzweig, Mark R. *Handbook of Development Economics*, chapter Labour markets in low-income countries. North-Holland, Amsterdam, 1988.

Stock, James H. and Jonathan H. Wright and Motohiro Yogo. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, 20(4):518-529, 2002.

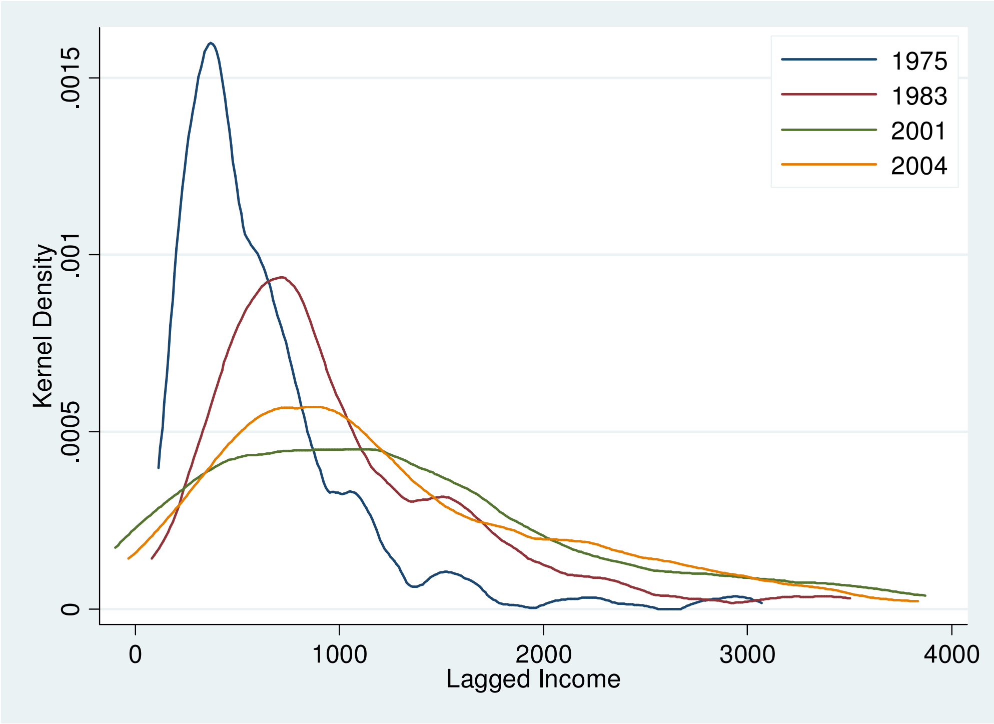
Swamy, Anand V. A simple test of the nutrition-based efficiency wage model. *Journal of Development Economics*, 53:85-98, 1997.

Walker, Thomas S. and James G. Ryan. *Village and Household Economies in India's Semi-Arid Tropics*. Johns Hopkins Press, Baltimore, 1990.

Wooldridge, Jeffrey M. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. Institute for Fiscal Studies Working Paper CWP11/02, 2000.

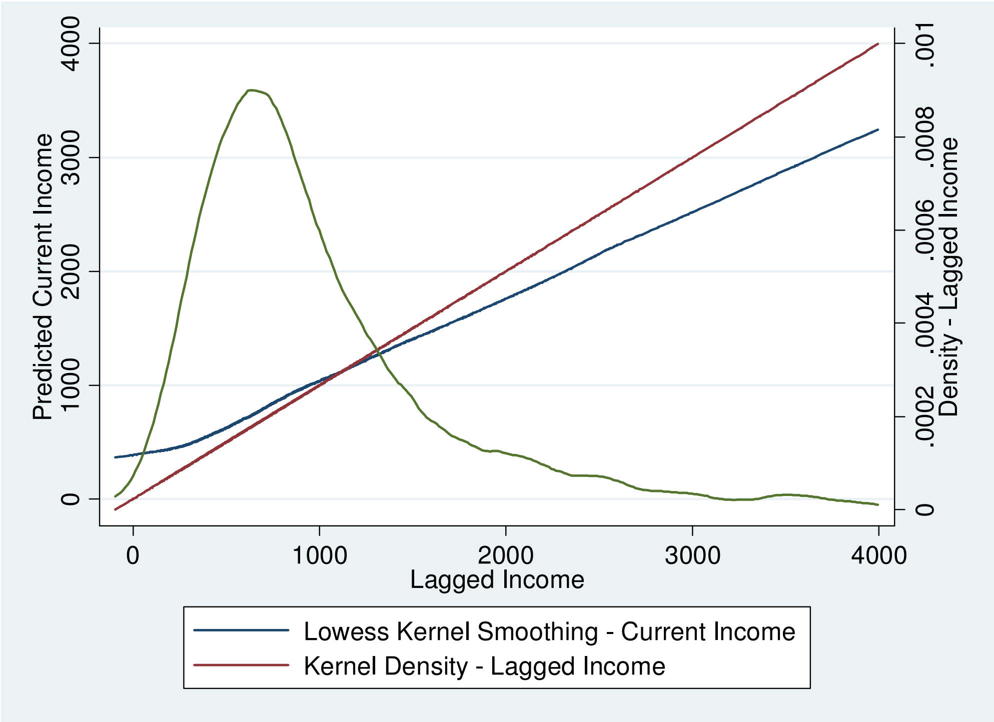
Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2002.

Graph [1]: Evolution of Income Distribution Over 30-Year Period
Kernel Densities of Income for Years 1975, 1983, 2001 and 2004



Note: Kernel smoothing densities for specific years. For ease of presentation income plotted are restricted to values -100 and 4000 Rupees in 1975 prices.

Graph [2]: Bivariate Lowess Estimates of Current Income on Lagged Income, and Kernel Densities of Lagged Income

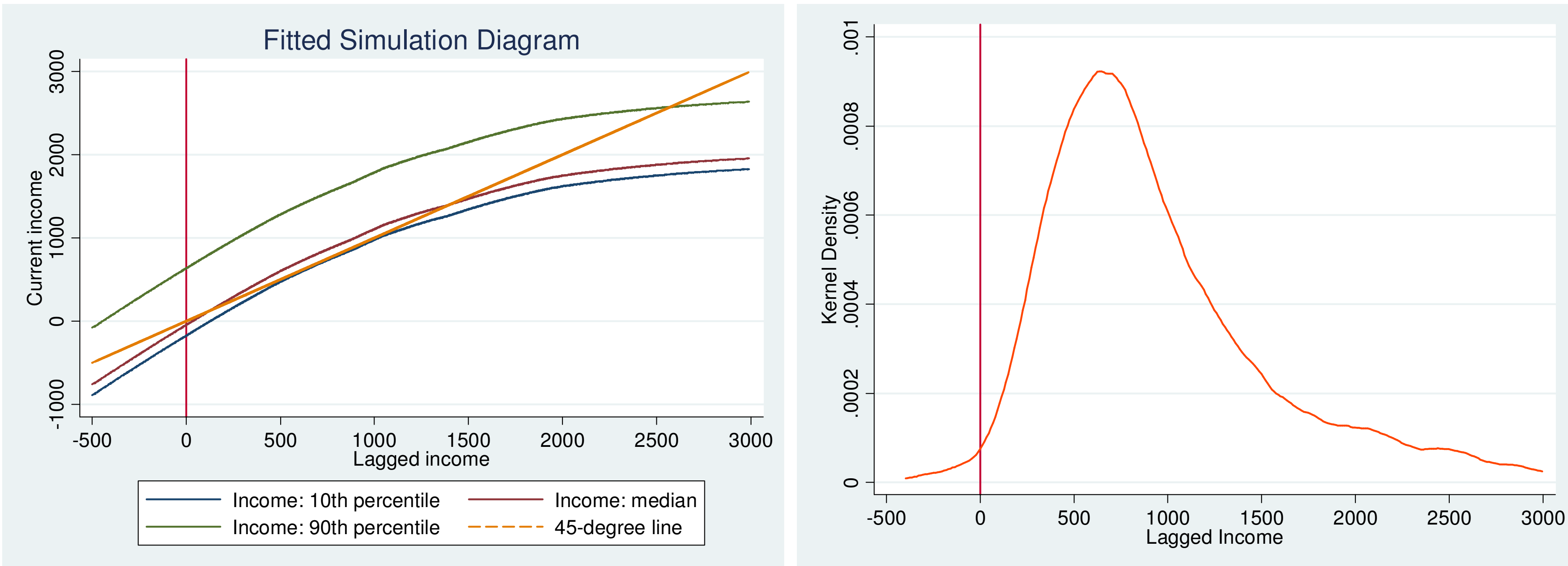


Note: 45-degree line indicates locus where current income equals lagged income. Data used for the graphs is restricted to the years used in the parametric analysis, namely 1977-1983, and 2003-2004. For ease of presentation income plotted are restricted to values -100 and 4000 Rupees in 1975 prices. Parametric analysis in the paper uses all income values.

Graph [3]: Simulated Income Dynamics for Median, 10th percentile and 90th percentile of fixed effects
Model Specification: IV GMM Estimates of Quadratic Model with Year Dummies,

Panel A – Simulated Dynamics

Panel B – Kernel Density of Lagged Income



Note: Simulations of income dynamics have been computed using the estimated parameter of the polynomial of lagged income. Plotted simulation based on IV GMM estimates for model specification with quadratic polynomial and year dummies. Computed fixed effects act as shifters of the locus of polynomial. Simulation curves are depicted for a realistic range of lagged-income. Intersections between the dynamic trajectories and the 45-degree line indicate a potential equilibrium. Vertical line indicates zero income. Panel B kernel densities of lagged income does not show 104 observations with values of lagged income above 3000 Rupees, as well as 10 observations with values below -500. In both cases, the values not shown are otherwise included in all regressions.

Table 1: Tracking and attrition in the 2001-2004 survey

Status by 2004-05	Full sample of individuals included in 1975-1984 (VLS1) with tracking information in 2005	Of which: Included in the 2001 survey, i.e. in the village and in the sample in 2001	Of which: Not included in the VLS2, 2001 survey
Dead in 2005?	432	24	408
Migrated in 2005?	675	45	630
In village in 2005?	857	581	276
No information in 2005?	34	4	30
Total	1998	654	1344

Note: based on attempts to track 1998 individuals included at some point between 1975-84 in the original households of the 1975-84 sample. Not including servants.

Table 2. Mean and Standard Deviation of Income, by Survey Year

	VLS 1									VLS 2				*Total*
	1975	*1976*	*1977*	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*	*2001*	*2002*	*2003*	*2004*	
Income	640.3	864.5	982.1	1019.7	1035.8	956.9	1018.5	1159.3	1044.4	1975.2	2353.0	1310.2	1257.8	1199.6
	(567.1)	(744.1)	(804.2)	(764.9)	(790.0)	(790.8)	(806.0)	(891.6)	(763.5)	(7,378.1)	(11,408.4)	(1,850.0)	(2,351.0)	(3,904.8)
Observations	255	258	262	262	263	273	273	274	274	261	260	258	255	3428

Note: Income measured in 1975 rupees per adult equivalent per year. Standard deviations appear in parentheses beneath mean values of continuous variables.

Table 3. First Stage Results – Impact of Rainfall on Lagged Income

Dependent variable:	ΔY_{t-1}	ΔY_{t-1}	ΔY_{t-1}^{2}	ΔY_{t-1}^{2}	ΔY_{t-1}^{3}	ΔY_{t-1}^{3}
	(1)	(2)	(3)	(4)	(5)	(6)
Rainfall shock, t-1	52.86***	81.2200	66,382.51	725,763.91	126370329.2	5.80E+09
	(18.290)	(73.600)	(9.81E+04)	(4.64E+05)	(7.94E+08)	(3.78E+09)
Rainfall shock * land operated (ha), t-1	5.12***	3.76**	30,448.00***	26,181.76**	154787235.39***	143272054.90**
	(1.720)	(1.690)	(9.14E+03)	(1.06E+04)	(5.62E+07)	(6.56E+07)
Rainfall shock * children 0-8 in HH, t-1	-19.83***	-31.10***	-73,784.72**	-119,976.96***	-4.12e+08*	-6.44e+08***
	(6.150)	(5.340)	(3.49E+04)	(2.97E+04)	(2.42E+08)	(2.04E+08)
ΔYear: 1977		184.23**		-58,039.18		-3.67E+09
		(73.770)		(4.67E+05)		(3.80E+09)
ΔYear: 1978		79.3500		-438,700.81		-5.59E+09
		(97.570)		(5.92E+05)		(4.54E+09)
ΔYear: 1979		-23.9900		-883,580.64		-7.74E+09
		(108.860)		(6.60E+05)		(5.23E+09)
ΔYear: 1980		-35.7500		-602,736.51		-5.22E+09
		(81.230)		(5.00E+05)		(3.90E+09)
ΔYear: 1981		-115.8400		-715,383.33		-5.54e+09*
		(71.960)		(4.38E+05)		(3.34E+09)
ΔYear: 1982		4.0500		-674,154.43		-5.95E+09
		(96.550)		(5.90E+05)		(4.72E+09)
ΔYear: 1983		93.8100		-81,396.43		-2.69E+09
		(58.050)		(3.62E+05)		(2.89E+09)
ΔYear: 2003		72.9800		-123,823.09		-9.47E+08
		(105.660)		(5.78E+05)		(4.43E+09)
ΔYear: 2004		-689.96***		-3477212.11***		-2.25e+10**
		(151.820)		(1.23E+06)		(1.07E+10)
Observations	2248	2248	2248	2248	2248	2248
R-squared	0.0000	0.0600	0.0000	0.0200	0.0000	0.0100
F Test for joint significance of IV	11.27	11.47	3.97	5.48	3.18	3.55

Note: Standard errors robust to heteroskedasticity and household-level clustering appear in parentheses. Income measured as real Rupees per adult equivalent. Rainfall shock measured as deviation from mean rainfall across observed waves of the panel.

Table 4. Modelling Income Dynamics, TSLS GMM Estimates

Dependent variable: ΔYt	Linear Polynomial		Quadratic Polynomial		Cubic Polynomial	
	(1)	(2)	(3)	(4)	(5)	(6)
ΔY_{t-1}	-0.040221	0.5319124***	0.179322	1.4574380***	0.891272	2.1515322**
	(0.219)	(0.187)	(0.368)	(0.479)	(1.376)	(0.924)
ΔY_{t-1}^{2}			-0.000065	-0.0002763**	-0.000556	-0.000837
			(0.000)	(0.000)	(0.001)	(0.001)
ΔY_{t-1}^{3}					0.000000	0.000000
					(0.000)	(0.000)
ΔYear: 1977		52.7805		-27.4243		1.7347
		(47.369)		(64.855)		(68.988)
ΔYear: 1978		-53.9930		-81.8115		-44.8529
		(48.196)		(55.255)		(61.983)
ΔYear: 1979		4.8351		-33.5984		-41.0478
		(31.397)		(40.812)		(41.855)
ΔYear: 1980		-79.5996190*		-72.2559		-74.1045621*
		(42.504)		(45.081)		(44.390)
ΔYear: 1981		91.9102651**		138.9750006**		155.4561876*
		(43.697)		(58.681)		(82.135)
ΔYear: 1982		95.5066867***		69.4890		46.9323
		(32.619)		(43.444)		(70.288)
ΔYear: 1983		-178.9808074***		-194.0245059***		-162.7718092**
		(43.983)		(47.057)		(58.015)
ΔYear: 2003		-460.2581603***		-481.6388404***		-585.3554501**
		(122.901)		(138.147)		(202.049)
ΔYear: 2004		603.2967086***		322.8020		523.2767
		(190.981)		(238.326)		(403.580)
Observations	2248	2248	2248	2248	2248	2248
R-squared	0.035	-0.728	0.044	-1.131	-2.131	-2.59
Second-Stage Diagnostics						
Anderson-Rubin F stat	0.615	5.648	0.615	5.648	0.615	5.648
Prob > F	0.606	0.001	0.606	0.001	0.606	0.001
Hansen J Statistic (overidentification)	1.699	6.137	1.25	1.991	-	-
Prob > chi-squared	0.428	0.0465	0.264	0.158	-	-

Source: Analysis of ICRISAT VLS 1975-2004 data. Standard errors robust to heteroskedasticity and clustering within household-year cells appear in parentheses. Income measured in 1975 Rupees per adult equivalent. Anderson Rubin F statistic tests the null hypothesis that the endogenous regressors are jointly insignificant in structural equation. Hansen J Statistic, reported for overidentified models tests null hypothesis that the excluded instruments are uncorrelated with the structural equation error. Cragg-Donald F-Statistics reported includes Kleibergen-Paap rank correction. Cragg-Donald 'Weak IV' statistic tests for the null hypothesis that instruments are strongly correlated with the set of endogeneous variables.

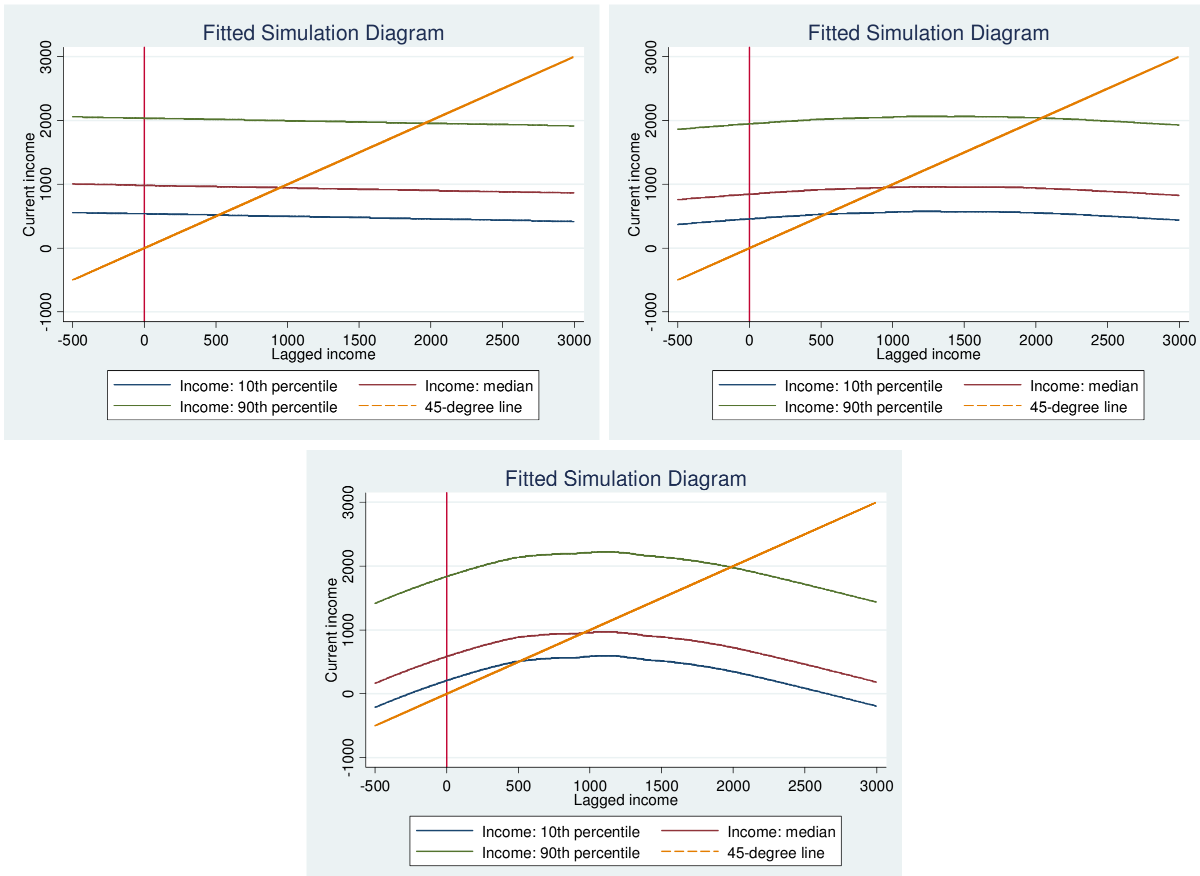
Table 5. Unpacking Household Fixed Effects - Correlates of HH Steady State Income

Dependent variable: HH Fixed Effect	Linear Polynomial		Quadratic Polynomial		Cubic Polynomial		Linear Polynomial		Quadratic Polynomial		Cubic Polynomial	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HH Head Education, 1975	246.36***	118.02***	276.14***	246.73***	546.83***	585.92***	153.80***	75.69***	175.69***	170.33***	350.25**	388.12**
	(78.08)	(34.42)	(88.50)	(83.04)	(185.17)	(203.52)	(55.20)	(23.86)	(61.61)	(56.58)	(129.98)	(143.43)
HH Head Age, 1975	7.58	2.42	7.88	3.83	11.31	8.94	4.98	1.30	5.05	1.68	5.30	2.82
	(6.41)	(3.06)	(6.93)	(5.57)	(12.54)	(12.82)	(4.75)	(2.38)	(5.09)	(4.27)	(9.02)	(9.43)
HH Head Sex, 1975	-282.58	-170.78	-283.79	-175.29	-544.82	-495.39	-170.77	-116.58	-163.28	-83.77	-332.96	-285.58
	(312.02)	(165.44)	(315.90)	(185.73)	(573.54)	(508.95)	(334.71)	(181.91)	(342.00)	(219.82)	(604.79)	(551.13)
HH Size, 1975	14.30	18.92	10.80	3.66	19.27	11.70	-50.85	-11.63	-59.69	-49.91	-112.71	-120.26
	(41.22)	(20.65)	(44.00)	(33.96)	(84.86)	(86.37)	(35.80)	(18.89)	(39.69)	(36.57)	(74.57)	(82.20)
Nr of Members ages 9-14, 1975	-89.80	-42.28	-95.37	-66.31	-233.10	-233.54	-21.88	-11.74	-21.53	-10.13	-84.53	-83.48
	(78.01)	(36.30)	(81.57)	(55.77)	(154.17)	(148.28)	(53.14)	(27.37)	(55.35)	(43.30)	(102.38)	(102.19)
Nr of Members ages 0-8, 1975	-110.84*	-70.25***	-111.52*	-73.33	-155.26	-132.62	-11.51	-24.09	-3.94	8.46	49.51	72.61
	(56.02)	(23.83)	(60.57)	(47.78)	(128.38)	(133.50)	(41.11)	(18.73)	(44.72)	(38.84)	(107.92)	(118.63)
Area Owned by HH (in Ha), 1975							59.40***	28.67***	64.05***	48.65***	113.61***	112.66***
							(8.48)	(4.21)	(8.66)	(9.79)	(16.25)	(18.32)
Value of HH Assets (in Ru), 1975							0.01***	0.00**	0.01***	0.01***	0.03***	0.03***
							(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Low Caste Dummy	-369.76***	-169.44***	-353.29***	-103.01	-459.95**	-265.97	-165.23	-77.39	-130.91	66.17	-13.04	185.37
	(104.06)	(53.48)	(111.57)	(101.97)	(214.97)	(222.30)	(103.49)	(49.14)	(111.97)	(103.30)	(216.67)	(229.13)
Village 2 Dummy	-106.04	-97.40	-104.35	-92.37	-222.21	-224.16	38.53	-33.77	53.22	27.57	105.68	108.51
	(154.70)	(65.35)	(153.37)	(87.22)	(265.42)	(239.91)	(154.22)	(65.91)	(151.60)	(81.37)	(251.54)	(214.30)
Village 3 Dummy	-263.83	-164.05**	-280.17	-235.89*	-599.90*	-619.39**	-148.93	-115.86*	-154.29	-139.98	-319.42	-332.37
	(170.21)	(76.23)	(174.57)	(118.64)	(294.11)	(274.35)	(141.15)	(63.43)	(143.25)	(100.35)	(232.85)	(216.16)
Village 4 Dummy	-250.42	-186.45***	-280.42*	-313.79***	-543.10*	-627.95**	-290.62*	-208.97***	-322.93**	-345.95***	-593.98**	-674.60**
	(152.00)	(64.35)	(152.90)	(87.12)	(283.78)	(261.88)	(142.41)	(62.60)	(142.98)	(89.29)	(270.15)	(255.71)
Village 5 Dummy	194.40	35.75	215.46	128.33	187.48	138.30	323.20*	91.47	356.10**	235.41*	487.63*	443.82*
	(196.18)	(83.84)	(205.46)	(145.05)	(301.39)	(271.11)	(158.79)	(66.63)	(164.71)	(117.48)	(247.54)	(219.64)
Village 6 Dummy	-256.22	-165.33*	-311.38	-400.40**	-758.04*	-932.66**	-62.27	-84.37	-98.79	-238.41**	-281.39	-444.52
	(215.70)	(97.18)	(222.92)	(152.23)	(413.67)	(402.93)	(164.56)	(72.82)	(166.67)	(112.90)	(320.36)	(317.25)
Village 7 Dummy	972.59*	537.80**	847.75	46.83	975.37	259.59	887.93*	500.34**	755.53	-23.36	785.03	66.67
	(504.19)	(241.29)	(520.38)	(339.16)	(962.69)	(903.09)	(456.94)	(227.52)	(467.75)	(306.05)	(854.93)	(798.29)
Observations	219	219	219	219	219	219	219	219	219	219	219	219
R-squared	0.281	0.278	0.281	0.262	0.26	0.25	0.477	0.448	0.481	0.438	0.492	0.475

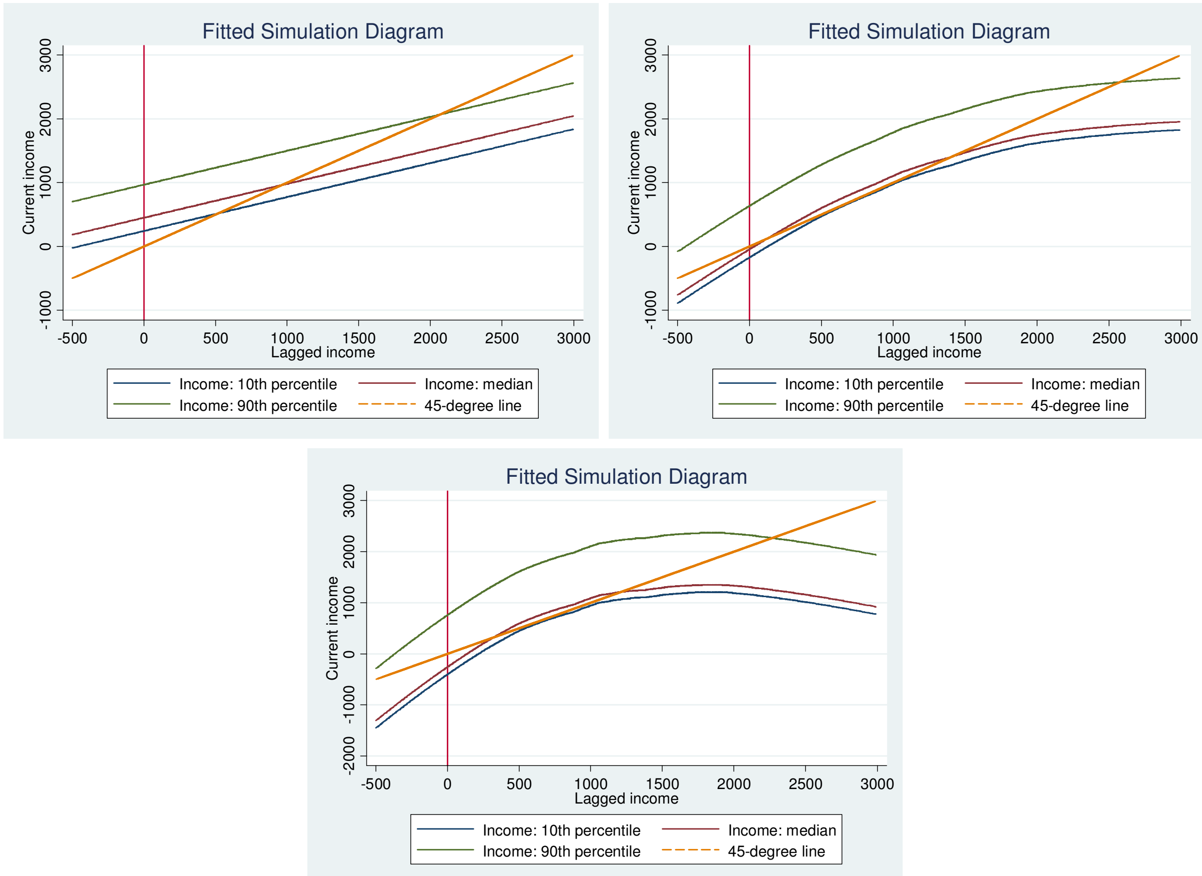
Note: Robust standard errors eported in parentheses. We restrict the sample size to the number of households for which we can compute income fixed effects. Fixed effects used in columns (1) to (6) and (7) to (12) were obtained using parameter estimates corresponding to columns (1) to (6) model specifications reported in Table 3.

**Graph [A1]: Simulated Income Dynamics for Median, 10th percentile and 90th percentile of fixed effects
All Model Specifications: Linear, Quadratic and Cubic**

Panel A – Linear, Quadratic and Cubic



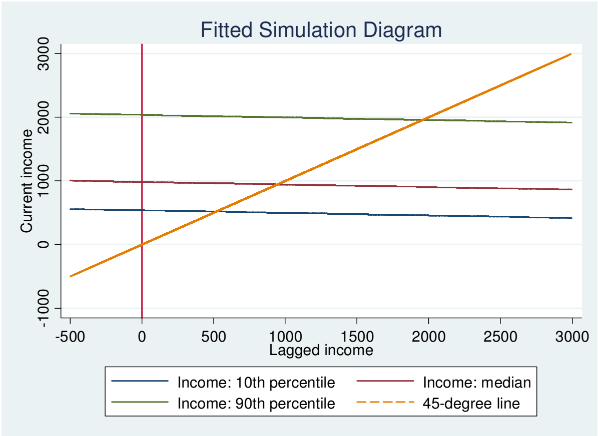
Panel B – Linear, Quadratic and Cubic with Year Dummies

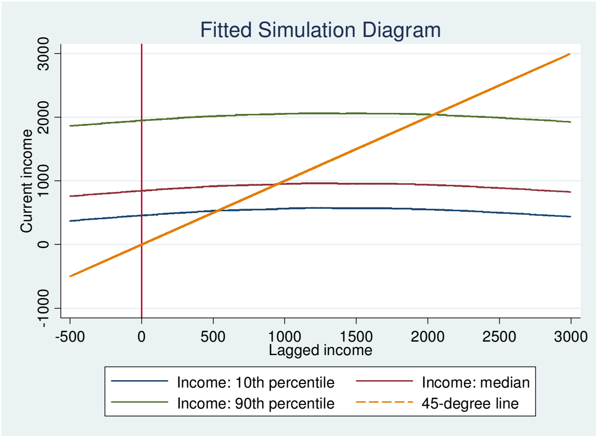


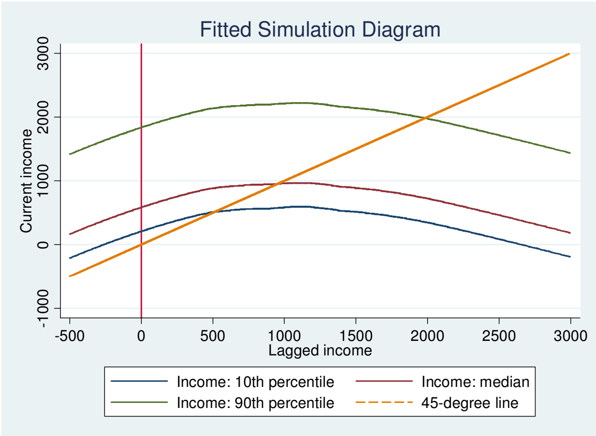
Note: Simulations of income dynamics have been computed using the estimated parameter of the polynomial of lagged income. Computed fixed effects act as shifters of the locus of polynomial. Simulation curves are depicted for a realistic range of lagged-income. Intersections between the dynamic trajectories and the 45-degree line indicate a potential equilibrium. Vertical line indicates zero income.

Graph [A1]: Simulated Income Dynamics for Median, 10th percentile and 90th percentile of fixed effects All Model Specifications: Linear, Quadratic and Cubic

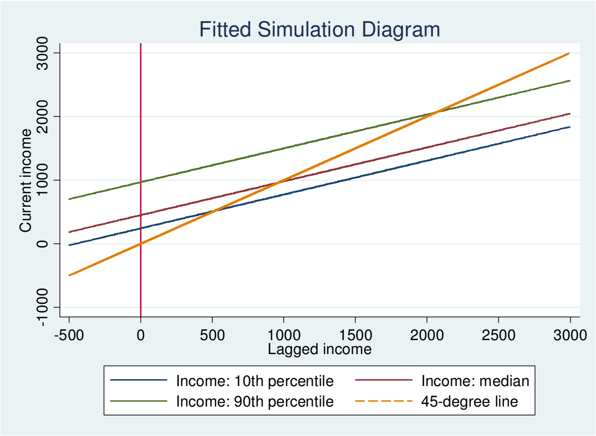
Panel A – Linear, Quadratic and Cubic

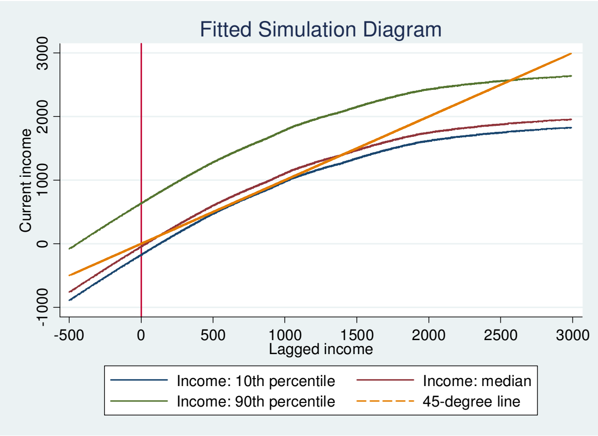


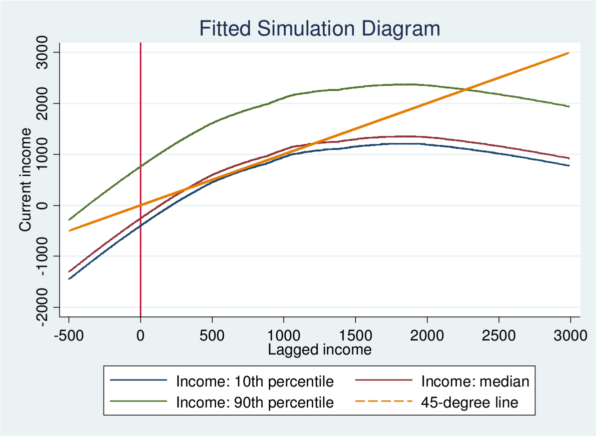




Panel B – Linear, Quadratic and Cubic with Year Dummies





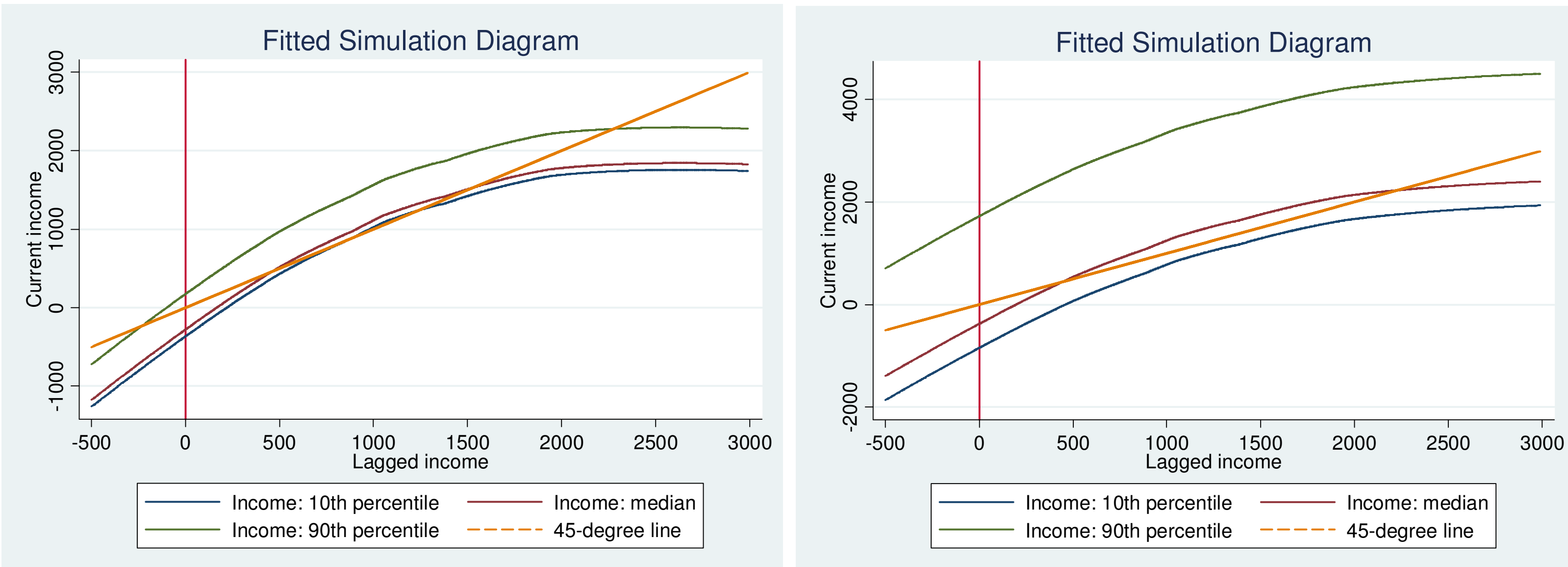


Note: Simulations of income dynamics have been computed using the estimated parameter of the polynomial of lagged income. Computed fixed effects act as shifters of the locus of polynomial. Simulation curves are depicted for a realistic range of lagged-income. Intersections between the dynamic trajectories and the 45-degree line indicate a potential equilibrium. Vertical line indicates zero income.

Graph [A3]: VLS1 and VLS2 Dynamics – Simulated Income Dynamics with Median, 10th percentile and 90th percentile of fixed effects
Model Specification: Quadratic Polynomial Model with Year Dummies by Panel Period

Panel A – VLS1 (1975-1983)

Panel B – VLS1 (2001-2004)



Note: Simulations of income dynamics have been computed using the estimated parameter of the polynomial of lagged income. Computed fixed effects act as shifters of the locus of polynomial. Simulation curves are depicted for a realistic range of lagged-income. Intersections between the dynamic trajectories and the 45-degree line indicate a potential equilibrium. Vertical line indicates zero income. Estimation methodology drops two years from each panel period, 1975-74 and 2001-2002 in VLS1 and VLS2 respectively.

Table A1. Robustness Checks - TSLS LIML and TSLS Fuller

Dependent variable: ΔYt	LIML TSLS Estimates			Fuller TSLS Estimates		
	Linear Polynomial	Quadratic Polynomial	Cubic Polynomial	Linear Polynomial	Quadratic Polynomial	Cubic Polynomial
	(1)	(2)	(3)	(4)	(5)	(6)
ΔY_{t-1}	0.5452230*	1.3747701*	2.1515322**	0.400700	1.0122233**	1.2135817***
	(0.331)	(0.708)	(0.924)	(0.268)	(0.423)	(0.395)
ΔY_{t-1}^{2}		-0.0002157	-0.000837		-0.000154	-0.0003472**
		(0.000)	(0.001)		(0.000)	(0.000)
ΔY_{t-1}^{3}			0.000000			0.000000
			(0.000)			(0.000)
ΔYear: 1977	13.9849069	-47.7631	1.7347	46.2397	-1.9216	22.4936
	(-70.758)	(77.090)	(68.988)	(58.077)	(58.182)	(50.959)
ΔYear: 1978	-43.1131818	-81.7873	-44.8529	-23.7125	-53.7785	-36.8911
	(-65.224)	(63.281)	(61.983)	(56.029)	(52.500)	(49.302)
ΔYear: 1979	-6.3901965	-35.8641	-41.0478	-0.7069	-22.5017	-22.5806
	(-36.227)	(46.034)	(41.855)	(33.247)	(38.195)	(36.012)
ΔYear: 1980	-81.0020274*	-78.8992155*	-74.1045621*	-78.5088478*	-77.3101444*	-75.0904329*
	(-45.205)	(46.785)	(44.390)	(43.080)	(44.237)	(42.653)
ΔYear: 1981	96.6846359*	126.9684644*	155.4561876*	87.0024367*	109.8685011*	115.9691290*
	(-51.380)	(67.753)	(82.135)	(46.189)	(56.448)	(59.367)
ΔYear: 1982	98.9536262**	75.4202	46.9323	109.2556494***	91.1450854**	85.2590449*
	(-41.694)	(50.390)	(70.288)	(36.976)	(41.410)	(44.673)
ΔYear: 1983	-189.8215336***	-208.1811944***	-162.7718092***	-169.8225183***	-185.4216336***	-165.6450532***
	(-57.672)	(51.149)	(58.015)	(49.766)	(44.471)	(43.423)
ΔYear: 2003	-466.4809810***	-478.5441391***	-585.3554501***	-448.3862661***	-459.2445232***	-486.9541585***
	(-138.666)	(149.300)	(202.049)	(127.306)	(131.663)	(140.723)
ΔYear: 2004	549.4638479**	437.2782	523.2767	456.0534739*	387.2351795*	385.8545412*
	(-270.853)	(273.683)	(403.580)	(233.061)	(228.753)	(227.718)
Observations	2248	2248	2248	2248	2248	2248
R-squared	-0.7550	-0.9340	-2.5900	-0.4830	-0.5430	-0.6810
First-Stage Diagnostics						
Number of Instruments	3	3	3	3	3	3
Cragg-Donald F-Statistic	11.47	2.48	0.49	11.47	2.48	0.49
Second-Stage Diagnostics						
Anderson-Rubin F stat	5.648	5.648	5.648	5.648	5.648	5.648
Prob > F	0.001	0.001	0.001	0.001	0.001	0.001
Hansen J Statistic (overidentification)	5.427	1.761	-	1.360	0.520	-
Prob > chi-squared	0.066	0.185	-	0.507	0.471	-

Note: Standard errors robust to heteroskedasticity and household-level clustering appear in parentheses. Income measured in 1975 Rupees per adult equivalent. Anderson Rubin F statistic tests the null hypothesis that the endogenous regressors are jointly insignificant in structural equation. Hansen J Statistic, reported for overidentified models tests null hypothesis that the excluded instruments are uncorrelated with the structural equation error. Cragg-Donald F-Statistics reported includes Kleibergen-Paap rank correction. Cragg-Donald 'Weak IV' statistic tests for the null hypothesis that instruments are strongly correlated with the set of endogenous variables.

Table A2. Robustness Checks - Parsimonious IV Set: Interactions (Rainfall x Land) and (Rainfall x Kids)

Dependent variable: ΔYt	Quadratic Polynomial Model	
	GMM or LIML Estimators	Fuller Estimator
	(1)	(2)
ΔY_{t-1}	1.7268980***	1.1053874***
	(0.581)	(0.324)
ΔY_{t-1}^{2}	-0.0003471**	-0.0002045***
	(0.000)	(0.000)
ΔYear: 1977	-51.1554	5.9872
	(73.194)	(54.238)
ΔYear: 1978	-84.7624	-49.3782
	(62.102)	(50.234)
ΔYear: 1979	-47.7931	-25.4332
	(45.833)	(37.261)
ΔYear: 1980	-74.9713	-75.0995174*
	(47.679)	(43.800)
ΔYear: 1981	135.1463401*	110.2271740**
	(69.665)	(55.891)
ΔYear: 1982	72.0140	92.7845468**
	(57.392)	(42.975)
ΔYear: 1983	-198.1413051***	-175.8830890***
	(51.073)	(42.719)
ΔYear: 2003	-466.6884612***	-449.5562886***
	(180.862)	(135.899)
ΔYear: 2004	269.7582	296.6419
	(289.062)	(214.984)
Observations	2248	2248
R-squared	-1.7200	-0.6300
First-Stage Diagnostics		
F-Statistic - ΔY_{t-1}	15.08	15.08
F-Statistic - ΔY_{t-1}^{2}	6.98	6.98
Number of Instruments	2	2
Cragg-Donald F-Statistic	3.35	3.35
Second-Stage Diagnostics		
Anderson-Rubin F stat	0.166	8.436
Prob > F	0.847	0.000

Note: Standard errors robust to heteroskedasticity and household-level clustering appear in parentheses. Income measured in 1975 Rupees per adult equivalent. Anderson Rubin F statistic tests the null hypothesis that the endogenous regressors are jointly insignificant in structural equation. Hansen J Statistic not defined since both models are only just-identified Cragg-Donald F-Statistics reported includes Kleibergen-Paap rank correction. Cragg-Donald 'Weak IV' statistic tests for the null hypothesis that instruments are strongly correlated with the set of endogenous variables. Parsimonious instruments are as follows: (Rainfall x Land Operated (Ha)) and (Rainfall x Nr Kids 0-8).

Table A3. Robustness Checks - Stability of Parameters Between VLS1 and VLS2, TSLS GMM Estimator

Dependent variable: ΔYt	Quadratic Polynomial Model			
	VLS1		VLS2	
	(1)	(2)	(3)	(4)
ΔY_{t-1}	1.1690928***	1.8247166***	2.014811	2.063861
	(0.298)	(0.456)	(2.413)	(2.265)
ΔY_{t-1}^2	-0.0002431**	-0.0003925***	-0.000274	-0.000398
	(0.000)	(0.000)	(0.000)	(0.000)
ΔYear: 1977		-66.8549		
		(49.887)		
ΔYear: 1978		-79.1821		
		(53.281)		
ΔYear: 1979		-53.1049		
		(38.517)		
ΔYear: 1980		-85.3914316*		
		(44.409)		
ΔYear: 1981		167.8761505***		
		(53.963)		
ΔYear: 1982		54.2389		
		(42.624)		
ΔYear: 1983		-198.5292210***		
		(42.751)		
ΔYear: 2003				-485.4916698*
				(251.244)
ΔYear: 2004				344.6865
				(585.742)
Observations	1836	1836	412	412
R-squared	-0.7050	-1.4630	-1.9170	-2.5500
First-Stage Diagnostics				
Number of Instruments	3	3	3	3
F-Statistic - ΔY_{t-1}	57.37	13.14	3.52	1.8
F-Statistic - ΔY_{t-1}^2	20.21	8.9	1.94	1.69
Cragg-Donald F-Statistic	5.05	2.16	0.21	0.68
Second-Stage Diagnostics				
Anderson-Rubin F stat	21.050	12.420	3.008	1.612
Prob > F	0.000	0.000	0.031	0.187
Hansen J Statistic (overidentification)	5.591	4.898	1.121	0.0173
Prob > chi-squared	0.018	0.027	0.29	0.895

Note: Standard errors robust to heteroskedasticity and household-level clustering appear in parentheses. Income measured in 1975 Rupees per adult equivalent. Anderson Rubin F statistic tests the null hypothesis that the endogenous regressors are jointly insignificant in structural equation. Hansen J Statistic, reported for overidentified models tests null hypothesis that the excluded instruments are uncorrelated with the structural equation error. Cragg-Donald 'Weak IV' statistic tests for the null hypothesis that instruments are strongly correlated with the set of endogenous variables. Estimation methodology drops two years from each panel period, 1975-74 and 2001-2002 in VLS1 and VLS2 respectively.