The pace of poverty reduction

*A fractional response approach**

Richard Bluhm† Denis de Crombrugghe‡ Adam Szirmai§

March 2015

Working Paper

Abstract

The pace of poverty reduction through growth vs. redistribution is at the heart of current debates on equitable development. In this paper, we argue that empirical poverty decompositions should build in the inherent boundedness of the poverty headcount ratio directly. As a solution, we propose a fractional response approach to estimating poverty decompositions, and present extensions dealing with unobserved heterogeneity, measurement error and unbalancedness. Using a large new data set, we estimate income and inequality (semi-)elasticities of poverty for the 2$ a day and 1.25$ a day poverty lines. The models fit the data remarkably well over the entire data range. We highlight the relevance of focusing on semi-elasticities for policy purposes and, building on the improved accuracy of the fractional response results, we present poverty projections from 2010 through 2030. Finally, we discuss some implications of these results for the post-2015 development agenda.

Keywords: poverty, inequality, fractional response models, income growth

JEL Classification: I32, C25, O10, O15

*This paper has been presented at the Center for Global Development (Washington 2014), IARIW (Rotterdam 2014), International Conference of the Courant Research Center (Goettingen 2014), an OECD conference (Paris 2013), a research workshop in panel data methods (Mainz 2013) and several workshops in Maastricht. We have greatly benefited from discussions with several participants. In particular, we would like to thank Jeffrey Wooldridge, Laurence Chandy, Charles Kenny, Stephan Klasen, Christophe Muller, Florent Bresson, Nicolas Meisel, Thomas Roca, Bart Verspagen and Kaj Thomsson. We gratefully acknowledge financial support from the Agence Française de Développement (AFD). All remaining errors are unfortunately ours.

†*Corresponding author.* Leibniz University Hanover, Institute of Macroeconomics, and Maastricht University/ UNU-MERIT, e-mail:*bluhm@mak.uni-hannover.de*

‡Maastricht University, e-mail:*d.decrombrugghe@maastrichtuniversity.nl*

§UNU-MERIT, Maastricht University, e-mail: *szirmai@merit.unu.edu*

1 Introduction

The pace of poverty reduction through economic growth is at the heart of ongoing debates on inclusive growth and equitable development. Given the salience of this issue to both policy makers and academics, an increasingly large literature analyzes the impact of changes in incomes and inequality on poverty, as well as their respective contributions towards poverty reduction in the past decades (Ravallion and Chen, 1997; Besley and Burgess, 2003; Bourguignon, 2003; Kraay, 2006; Kalwij and Verschoor, 2007; Bresson, 2009; Chambers and Dhongde, 2011). Collectively, these studies established not only that income growth is crucial to achieving sustained decreases in poverty but also that the benefits of income growth strongly depend on the initial levels of income and inequality. In fact, poverty is linked to income and distribution through a decomposition identity that should guide empirical studies of poverty (Datt and Ravallion, 1992; Kakwani, 1993; Bourguignon, 2003). In this paper, we revisit and extend this literature in several ways.

First, and most fundamentally, we argue that the previous literature based on linear models of poverty change ignores the bounded nature of the poverty headcount ratio and either disregards or awkwardly reintroduces the information provided by income and inequality levels. This leads to poorly fitting approximations of the underlying decomposition identity and estimates of the relevant elasticities that can take on implausible values. Instead, we propose that empirical models of poverty should capture the boundedness of the poverty headcount ratio directly, and thus build in the inherent non-linearity of the relationship and the non-constancy of the elasticities. To this end, we derive fractional response models of the poverty headcount ratio. We present extensions of these models that deal with unobserved heterogeneity, measurement error and unbalanced panel data (see Papke and Wooldridge, 1996, 2008; Wooldridge, 2010a). A key result is the greatly improved accuracy of the poverty decomposition. Our specifications fit the data remarkably well and predict the observed poverty headcount ratio with less than 2.5 percentage points error on average. For comparison, we also reproduce the traditional approach and highlight some of its empirical shortcomings.

Second, using a new data set of 809 nationally representative surveys covering 124 countries from 1981 to 2010, we estimate income and inequality elasticities of the poverty headcount ratio for different regions and time periods. Our findings generally confirm the result of previous studies that the average income elasticity of poverty is around two. However, in order to understand the speed of poverty reduction across different regions and time periods, we are particularly interested in reliable estimates of values other than the overall cross-country mean. Our method provides considerably more precise regional and temporal estimates of the income and inequality elasticities of poverty that often contradict earlier studies. For example, we find universally higher income elasticities in Latin America and Eastern Europe and Central Asia but lower income elasticities in both South Asia and Sub-Saharan Africa than reported earlier (Kalwij and Verschoor, 2007). Furthermore, since income or inequality elasticities of poverty are concepts of relative change, they may give the misleading impression that richer countries are becoming ever better at reducing poverty even though the underlying absolute changes are small. Hence, we also stress the importance of semi-elasticities which capture the absolute change in poverty for a given rate of income growth or proportional change in the income distribution (Klasen and Misselhorn, 2008).

Last but not least, based on the much improved accuracy of our new estimates, we present projections of the poverty headcount ratio for the 2$ a day and 1.25$ a day poverty lines until 2030. We find that absolute poverty in Sub-Saharan Africa and, as a not too distant second, South Asia remains the primary development challenge of the twenty-first century. In all other regions, poverty is projected to nearly disappear or fall to much less than 10% of the population by 2030. These results are based on the assumption that per capita incomes and expenditures in each country continue to grow at the average pace of the last 15 years. While it is clearly far from guaranteed that this will hold true, other assumptions are easily evaluated using our proposed estimation framework. The projected poverty trends have direct bearing on the post-2015 development agenda and can be used to assess the viability of proposed development goals.

The remainder of this paper is organized as follows. Section 2 reviews how the existing literature models poverty, derives our approach, and discusses the econometrics of fractional response models. Section 3 briefly outlines the data used in this paper. Section 4 presents the estimation results, regional elasticities, and poverty projections until 2030. Section 5 concludes.

2 Modeling poverty and elasticities

2.1 Traditional approaches: linear models of poverty changes

Using micro-data, changes in poverty can be decomposed into changes in income and the income distribution up to some error (Datt and Ravallion, 1992; Kakwani, 1993). A key problem for cross-country studies of poverty is that we generally do not have access to micro data sets of incomes and/or consumption expenditures for all countries but usually estimate poverty based on grouped data.[1] To overcome this limitation, Bourguignon (2003) suggests approximating the entire income distribution of each country using a two-parameter log normal distribution – an approach that is theoretically grounded[2], simple and popular but not without its critics (e.g. Bresson, 2009).

Bourguignon assumes that income, y_t, is a log normal random variable, such that $\ln y_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$, and mean income can be written as $\bar{y}_t = \mathrm{E}[y_t] = \exp(\mu_t + \sigma_t^2/2)$. Then the poverty headcount ratio (henceforth, poverty headcount) at time t may be defined as

$$H_t = H(\bar{y}_t/z, \sigma_t) = \Phi\left(\frac{-\ln(\bar{y}_t/z)}{\sigma_t} + \frac{1}{2}\sigma_t\right) \equiv \Pr[y_t \leq z] \tag{1}$$

where $\Phi(\cdot)$ denotes the standard normal cdf, inequality is measured as the standard deviation (σ_t) of log income, and $\bar{y}_t/z$ is the (relative) distance of mean income ($\bar{y}_t$) to the (fixed) poverty line (z) – which we interpret as a 'shortfall' when $\bar{y}_t < z$ and 'affluence' when $\bar{y}_t > z$.

[1]There have been some attempts to either collect all the available primary data or to estimate gaps in survey coverage with the help of national accounts. Milanovic (2002) compiles a global data set of household level data to study the evolution of inequality, Sala-i-Martin (2006) estimates a "world income distribution" via kernel density approximations based on grouped data, and Kraay (2006) fits three-parameter Lorenz curves to grouped data in order to estimate the entire income distribution for the country-poverty spells in his sample.

[2]Gibrat's law, for example, illustrates how the log normal distribution can arise from a sequence of stochastic income shocks $\ln y_t = \ln y_{t-1} + e_t$, so that e_t is a random transitory shock in log income and as t grows the distribution of e_t defines the distribution of $\ln y_t$. Battistin, Blundell, and Lewbel (2009) recently argued that this process is better thought of in terms of permanent income and suggest that consumption is closer to a log normal distribution.

Eq. (1) can be interpreted as the probability that, at a particular time t, an individual randomly drawn from the population is poor. This formulation gave rise to a large literature deriving the income and inequality elasticities of poverty analytically and estimating econometric models inspired by their analytic counterparts (e.g. Bourguignon, 2003; Kalwij and Verschoor, 2007; Klasen and Misselhorn, 2008). To summarize the essence of this approach note that, based on eq. (1), we can derive the income elasticity ($\varepsilon_t^{H\bar{y}} = \frac{\partial H_t}{\partial \bar{y}_t}\frac{\bar{y}_t}{H_t}$) and inequality elasticity ($\varepsilon_t^{H\sigma} = \frac{\partial H_t}{\partial \sigma_t}\frac{\sigma_t}{H_t}$) of the poverty headcount as

$$\varepsilon_t^{H\bar{y}} = -\frac{1}{\sigma_t}\lambda\left(\frac{-\ln(\bar{y}_t/z)}{\sigma_t} + \frac{1}{2}\sigma_t\right) \tag{2}$$

and

$$\varepsilon_t^{H\sigma} = \left(\frac{\ln(\bar{y}_t/z)}{\sigma_t} + \frac{1}{2}\sigma_t\right)\lambda\left(\frac{-\ln(\bar{y}_t/z)}{\sigma_t} + \frac{1}{2}\sigma_t\right) \tag{3}$$

where we define the inverse Mills ratio ($\lambda(x) \equiv \phi(x)/\Phi(x)$) as the ratio of the standard normal pdf to the standard normal cdf, and we require $H_t > 0$.

The decomposition of the poverty headcount is often written as

$$\frac{\mathrm{d}H_t}{H_t} \approx \varepsilon_t^{H\bar{y}}\frac{\mathrm{d}\bar{y}_t}{\bar{y}_t} + \varepsilon_t^{H\sigma}\frac{\mathrm{d}\sigma_t}{\sigma_t} \tag{4}$$

where $\mathrm{d}H_t/H_t$ is a small relative change in the poverty headcount, $\mathrm{d}\bar{y}_t/\bar{y}_t$ is a small relative change in mean incomes, and $\mathrm{d}\sigma_t/\sigma_t$ is a small relative change in the standard deviation of log incomes. The approximation follows from a linear Taylor expansion of H_t. Appendix A derives this result.[3]

Given log normality, the standard deviation is a monotone transformation of the Gini inequality coefficient, denoted G_t, and can be obtained via $\sigma_t = \sqrt{2}\Phi^{-1}(G_t/2+1/2)$.[4] So, eqs. (2) and (3) can be used to predict the elasticities directly using observed values of income and inequality. With some additional algebra, we can also derive an expression for the Gini elasticity and rewrite eq. (4) accordingly – see eq. (A-6) in Appendix A.

However, the assumption of log normality is only an approximation and unlikely to hold exactly. The key observation motivating the econometric models is that both elasticities depend only on the initial levels of mean income and inequality (when the poverty line is fixed). In order to allow for misspecification of the functional form, the authors rely on a linear approximation of these intrinsically non-linear functions. They capture the dependence on initial levels by interacting both mean income and inequality with the ratio of initial mean income to the poverty line and with initial inequality. This model is sometimes called the "improved standard model" (Bourguignon, 2003) and it is usually formulated in (annualized) differences:

$$\begin{aligned}\Delta \ln H_{it} = \;& \alpha + \beta_1 \Delta \ln \bar{y}_{it} + \beta_2 \Delta \ln \bar{y}_{it} \times \ln(\bar{y}_{i,t-1}/z) + \beta_3 \Delta \ln \bar{y}_{it} \times \ln G_{i,t-1} \\ & + \gamma_1 \Delta \ln G_{it} + \gamma_2 \Delta \ln G_{it} \times \ln(\bar{y}_{i,t-1}/z) + \gamma_3 \Delta \ln G_{it} \times \ln G_{i,t-1} + \epsilon_{it}\end{aligned} \tag{5}$$

where Δ is the difference operator, α is a linear time trend and ϵ_{it} is an error term. We also added country subscripts.

[3] Also see Datt and Ravallion (1992), Kakwani (1993) and Bourguignon (2003).

[4] This formulation is the inverse of $G_t = 2\Phi(\sigma_t/\sqrt{2}) - 1$ due to Aitchison and Brown (1957).

Suppose eq. (5) is estimated via OLS or GMM, then it is straightforward to see that the implied elasticities approximate eqs. (2) and (3). The estimated income elasticity is $\hat{\varepsilon}_{it}^{H\bar{y}} = \hat{\beta}_1 + \hat{\beta}_2 \ln(\bar{y}_{i,t-1}/z) + \hat{\beta}_3 \ln G_{i,t-1}$ and the estimated inequality elasticity is $\hat{\varepsilon}_{it}^{HG} = \hat{\gamma}_1 + \hat{\gamma}_2 \ln(\bar{y}_{i,t-1}/z) + \hat{\gamma}_3 \ln G_{i,t-1}$. Clearly, the elasticities depend on the initial levels of income and inequality. These two elasticities are sometimes referred to as the 'distribution-neutral' income elasticity and the 'growth-neutral' inequality elasticity. They identify the partial effect of changing either income or inequality, contrary to the simple observed elasticity that confounds both effects.[5] Yet, this approximation is extremely coarse and does not place any meaningful restrictions on the parameter space. Moreover, it is unclear which level equation this specification derives from.

In general, poverty elasticities can paint a distorted picture of poverty dynamics. The income elasticity, for example, gives the impression that richer countries become ever better at poverty reduction because a drop in the poverty headcount from 2% to 1% is treated just the same as a drop from 50% to 25%. Recognizing this shortcoming, Klasen and Misselhorn (2008) suggest to focus on absolute poverty changes instead. Removing the log from the headcount in eq. (5) turns it into a model of semi-elasticities and alters the interpretation.[6] The coefficients now measure the percentage point change in the population that is below the poverty line for a given rate of change in income or inequality. Likewise, eqs. (2) and (3) can be written as semi-elasticities by replacing the inverse Mills ratio with the standard normal pdf. Contrary to elasticities, the semi-elasticities approach zero as mean income becomes large. Klasen and Misselhorn (2008) also report that their models fit the data better and suggest that the specification in absolute changes captures more of the inherent non-linearity. Nevertheless, given the underlying identity, the fit typically found in the literature ($.55 \leq R^2 \leq .73$) is not particularly high.

There are advantages to relying on a linear framework other than mere simplicity, such as the well-known robustness properties of popular estimators. Nevertheless, a specification like eq. (5) suffers from several problems. First, it completely disregards the information provided by poverty levels, most likely introduces negative serial correlation, and compounds pre-existing measurement error.[7] Second, after differencing removes the time-constant unobserved effects, the added interaction terms reintroduce the unobserved effects present in the lagged levels. Hence, the coefficients are likely to be biased no matter if the model is estimated using Ordinary Least Squares (OLS), Instrumental Variables (IV) or Generalized Methods of Moments (GMM) if time invariant measurement differences exist.[8] *Third and most fundamentally*, the intrinsic non-linearity of the (semi-) elasticities is due to the bounded nature of the dependent variable which is not taken into account by linear specifications. The poverty headcount is a proportion and thus only takes on values in the unit interval ($H_{it} \in [0, 1]$). As a result, the linear estimates are

[5]For recent estimates of the simple empirical elasticity ($\mathrm{d}\ln H_{it}/\mathrm{d}\ln rgdpc_{it}$, where $rgdpc_{it}$ is real GDP per capita), see Ram (2013).

[6]The term semi-elasticity refers here to the quantity $\frac{\partial \ln y}{\partial x} = \frac{\partial y}{\partial x}\frac{1}{y}$ rather than the more usual $\frac{\partial y}{\partial \ln x} = \frac{\partial y}{\partial x} x$.

[7]If the errors in the original process (say, $y_{it} = \alpha + \beta(x_{it} + \nu_{it}) + \epsilon_{it}$, where ν_{it} is a mean zero process uncorrelated with x_{it}) are not autocorrelated, then differencing introduces correlation: $\Delta\epsilon_{it}$ has a term in common with $\Delta\epsilon_{i,t-1}$, so $\mathrm{E}[\Delta\epsilon_{it}\Delta\epsilon_{i,t-1}] = \mathrm{E}[-\epsilon_{i,t-1}^2] = -\sigma_\epsilon^2$. Furthermore, any attenuation bias is magnified by the first-difference transformation: $\mathrm{plim}(\hat{\beta}_{FD}) = \beta\sigma_{\Delta x}^2/(\sigma_{\Delta x}^2 + \sigma_{\Delta\nu}^2)$ where typically $\sigma_{\Delta\nu}^2 = 2\sigma_\nu^2$ but $\sigma_{\Delta x}^2 < 2\sigma_x^2$. Autocorrelation in the mismeasured variables further reduces the signal to noise ratio and increases the attenuation bias.

[8]Cf. Kalwij and Verschoor (2007), who first show a simpler linear model in differences to remove the unobserved effects and then estimate interaction models with unobserved effects reintroduced.

unlikely to work well for values distant from the mean and can easily take on implausible values (e.g. $\hat{\varepsilon}_{it}^{H\bar{y}} > 0$) for some combinations of income and inequality. Any version of the linear model will only poorly approximate the non-linear shape. Yet for these estimates to be relevant for any particular country or region, we are precisely interested in temporal and/or regional elasticities and not just the overall cross-country mean elasticities.

2.2 Alternative approaches: non-linear models of poverty

We propose modeling the conditional expectation of the poverty headcount using non-linear parametric fractional response models. To the best of our knowledge, this paper is the first to use such an approach for estimating poverty elasticities.

Some weaknesses of the linear specifications have been highlighted before. Chambers and Dhongde (2011) argue that the functional form is the source of the non-linearity. They estimate non-parametric models of the conditional mean of the headcount ($H_{it} = m(\bar{y}_{it}, G_{it}) + \epsilon_{it}$) and then obtain average elasticities. However, non-parametric techniques are often inefficient, cannot handle many covariates (due to the curse of dimensionality), and cannot easily deal with measurement error. It is generally difficult to test more involved hypotheses in a non-parametric framework. Nevertheless, we fully agree with Chambers and Dhongde (2011) that non-linear estimation of the conditional mean matters.

We follow Papke and Wooldridge (1996), who first suggested modeling proportions using fractional logit or fractional probit to estimate models of the form $\mathrm{E}[y_i|\mathbf{x}_i] = F(\mathbf{x}_i'\boldsymbol{\beta})$, where $y_i \in [0,1]$ and $F(\cdot)$ is the logistic or normal cdf. Applying this approach to our problem, we may approximate eq. (1) with

$$\mathrm{E}[H_{it}|\bar{y}_{it}, G_{it}] = \Phi(\alpha + \beta \ln \bar{y}_{it} + \gamma \ln G_{it}) \quad \text{for} \quad i = 1, \ldots, N; \; t = 1, \ldots, T \tag{6}$$

where $\ln z$ is absorbed into the constant and we also take the logarithm of the Gini (mostly for convenience). Naturally, we expect $\beta < 0$ and $\gamma > 0$. In motivating this model, we temporarily assume away all econometric complications such as unobserved heterogeneity, endogeneity and unbalancedness. These assumptions are relaxed in the next subsection.

Since $F(\cdot)$ is invertible, it can be used as a 'link function' in the spirit of the GLM literature (e.g. MacCullagh and Nelder, 1989). Thus, we may also write eq. (6) as $\Phi^{-1}(\mathrm{E}[H_{it}|\bar{y}_{it}, G_{it}]) = \alpha + \beta \ln \bar{y}_{it} + \gamma \ln G_{it}$. In other words, the inverse normal cdf linearizes the conditional mean. Figures B-1 and B-2 in Appendix B use this property to plot $\Phi^{-1}(H_{it})$ against $\ln \bar{y}_{it}$ and $\ln G_{it}$ for each region, including a regression line. The result is striking. This simple transformation is extremely successful in removing the intrinsic non-linearity of the poverty headcount.

It is now straightforward to define the estimated income elasticity as

$$\hat{\varepsilon}_{it}^{H\bar{y}} = \frac{\partial \hat{\mathrm{E}}[H_{it}|\bar{y}_{it}, G_{it}]}{\partial \bar{y}_{it}} \frac{\bar{y}_{it}}{\hat{\mathrm{E}}[H_{it}|\bar{y}_{it}, G_{it}]} = \hat{\beta} \times \lambda \left(\hat{\alpha} + \hat{\beta} \ln \bar{y}_{it} + \hat{\gamma} \ln G_{it} \right) \tag{7}$$

and the estimated Gini elasticity as

$$\hat{\varepsilon}_{it}^{HG} = \frac{\partial \hat{\mathrm{E}}[H_{it}|\bar{y}_{it}, G_{it}]}{\partial G_{it}} \frac{G_{it}}{\hat{\mathrm{E}}[H_{it}|\bar{y}_{it}, G_{it}]} = \hat{\gamma} \times \lambda \left(\hat{\alpha} + \hat{\beta} \ln \bar{y}_{it} + \hat{\gamma} \ln G_{it} \right). \tag{8}$$

Contrary to the linear model, the elasticities in eqs. (7) and (8) closely mimic the properties and structure of the analytical elasticities in eqs. (2) and (3). The non-linearity arises simply from the bounded functional form and is not artificially captured by interaction terms. However, it is important to note that we *do not require log normality* but implicitly assume that an unspecified two-parameter distribution sufficiently describes the poverty headcount up to statistical error. We simply derived a more natural model of the poverty headcount as a function of income and inequality which happens to look a lot like its theoretical counterpart.

This approach has several advantages. We directly take the information provided by poverty levels into account, the model predictions will strictly lie in the unit interval, and the same model is able to estimate both elasticities and semi-elasticities. As a result of respecting the bounded nature of the headcount, the elasticities will also approach zero when the inverse Mills ratio becomes vanishingly small and are likely to closely approximate values further away from the mean of the covariates. Hence, they share important properties of their theoretical counterparts based on the log normal assumption. Furthermore, we can directly predict the poverty headcount for interesting combinations of income and inequality rather than going the roundabout way of estimating the elasticities in one model and then projecting poverty separately.

2.3 Econometrics of fractional response models

Since there is no free lunch in econometrics, the apparent gains over the linear approach must come at a price. In non-linear models it is generally harder to deal with unobserved heterogeneity, measurement error and unbalancedness. While until a few years ago there was relatively little research on this issue, we can now draw on an increasingly well developed framework. Papke and Wooldridge (2008) extend fractional response models to balanced panels with unobserved heterogeneity and endogenous covariates, Wooldridge (2010a) develops the theory for unbalanced panels, and Wooldridge (2012) derives a general set-up for one-step estimators in non-linear models. Other contributions to the field are Loudermilk (2007), Cook, Kieschnick, and McCullough (2008), Ramalho, Ramalho, and Murteira (2011), and Tiwari and Palm (2011).

To simplify the exposition, we stack the coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ and covariates $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, \ldots, x_{it,k})'$. The ideal model we would like to estimate is

$$\mathrm{E}[H_{it}|\mathbf{x}_i, \mu_i] = \mathrm{E}[H_{it}|\mathbf{x}_{it}, \mu_i] = \Phi(\mathbf{x}_{it}'\boldsymbol{\beta} + \mu_i) \quad \text{for} \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{9}$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})$ are the covariates in all periods. We assume that the covariates are strictly exogenous conditionally on the unobserved effects (μ_i), and that the panel is balanced. We introduced unobserved country-level effects to capture time-persistent differences in measurement or deviations from a two-parameter distribution, which may be arbitrarily correlated with the elements in $\mathbf{x}_{it}$.

The key problem with such an approach is that the unobserved effects are not identified when T is fixed and $N \longrightarrow \infty$, leading to biased estimates of the parameter vector. This is the incidental parameters problem of Neyman and Scott (1948).[9] In addition, the partial effects needed for calculating the elasticities are not identified either. Papke

[9]This bias tends to become small as T gets large, but there are no benchmark simulations for the fractional probit case that we know of and in our case $\bar{T}$ is small. Papke and Wooldridge (2008) explain why replacing the standard normal cdf by the logistic cdf is not a good solution for this problem.

and Wooldridge (2008) suggest to solve this problem by imposing some structure on the correlation between the unobserved effects and the covariates using a device developed by Mundlak (1978) and Chamberlain (1984). This approach is generally known as correlated random effects (CRE). Concretely, we let

$$\mu_i | (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) \sim \mathcal{N}(\varphi + \bar{\mathbf{x}}_i' \boldsymbol{\theta}, \sigma_u^2) \tag{10}$$

where $\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$ is the time average of all the included time-varying regressors, $\mathbf{x}_{it}$ no longer contains a constant, and $u_i = \mu_i - \varphi - \bar{\mathbf{x}}_i' \boldsymbol{\theta}$ with $u_i | (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) \sim \mathcal{N}(0, \sigma_u^2)$. The covariates are still strictly exogenous conditionally on the unobserved effects. In linear models, this specification is equivalent to the traditional 'fixed effects' model and thus, in terms of accounting for unobserved effects, achieves the same aim as specifying a difference equation.

Plugging eq. (10) into eq. (9), we can rewrite our model of interest as

$$\mathrm{E}[H_{it} | \mathbf{x}_i, \mu_i] = \Phi(\varphi + \mathbf{x}_{it}' \boldsymbol{\beta} + \bar{\mathbf{x}}_i' \boldsymbol{\theta} + u_i) \tag{11}$$

$$\mathrm{E}[H_{it} | \mathbf{x}_i] = \mathrm{E}[\Phi(\varphi + \mathbf{x}_{it}' \boldsymbol{\beta} + \bar{\mathbf{x}}_i' \boldsymbol{\theta} + u_i) | \mathbf{x}_i] = \Phi\left(\varphi_u + \mathbf{x}_{it}' \boldsymbol{\beta}_u + \bar{\mathbf{x}}_i' \boldsymbol{\theta}_u\right) \tag{12}$$

where the subscript u denotes scaling of the coefficients by the factor $(1 + \sigma_u^2)^{-1/2}$. Going from eq. (11) to eq. (12) applies iterated expectations and the last equality follows from mixing (compounding) independent mean-zero normals.

If these assumptions hold, then the scaled coefficients and average partial effects (APEs) of all time-varying covariates are identified. However, survey-specific (non-classical) measurement error in income is likely to lead to overestimating the income elasticity in absolute value (Ravallion and Chen, 1997). In addition, classical measurement error may attenuate the income coefficient and thus work in the opposite direction. Suppose we do not observe true income but $\ln \bar{y}_{it} = \ln \bar{y}_{it}^* + \upsilon_{it}$, where $\ln \bar{y}_{it}^*$ is the true value of log mean income/ expenditure and υ_{it} is a composite error process with a classical and a non-classical component. We can view this as an omitted variable problem. Going back to the simplest model, we have $\mathrm{E}[H_{it} | \mathbf{x}_{1it}, \bar{y}_{it}^*, \mu_i] \neq \mathrm{E}[H_{it} | \mathbf{x}_{1it}, \bar{y}_{it}, \mu_i] = \Phi(\mathbf{x}_{1it}' \boldsymbol{\beta} + \psi(\ln \bar{y}_{it}^* + \upsilon_{it}) + \mu_i) = \Phi(\mathbf{x}_{1it}' \boldsymbol{\beta} + \psi \ln \bar{y}_{it} - \psi \upsilon_{it} + \mu_i)$, where $\mathbf{x}_{1it}$ is $\mathbf{x}_{it}$ without the mismeasured $\ln \bar{y}_{it}$, and υ_{it} is also potentially correlated with the time-constant unobserved effects ($\mathrm{cov}(\upsilon_{it}, \mu_i) \neq 0$). Inference using observed income will lead to underestimating or overestimating the effect depending on which type of error is stronger.

Building on Rivers and Vuong (1988) and the general result from Blundell and Powell (2004), Papke and Wooldridge (2008) suggest a simple two-step control function estimator for such endogeneity problems. Provided we have an $m \times 1$ vector of time-varying instruments ($\mathbf{z}_{it}$) that are relevant but not correlated with υ_{it}, we can estimate a log-linear model $\ln \bar{y}_{it} = \varphi_1 + \mathbf{x}_{1it}' \boldsymbol{\beta}_1 + \mathbf{z}_{it}' \gamma_1 + \bar{\mathbf{x}}_i' \boldsymbol{\theta}_1 + \nu_{it}$ in the first step and then obtain the residuals $\hat{\nu}_{it}$. It is important to note here that we redefine $\bar{\mathbf{x}}_i = (\bar{\mathbf{x}}_{1i}, \bar{\mathbf{z}}_i)$; that is, $\bar{\mathbf{x}}_i$ now contains the time averages of *all* strictly exogenous variables including the instruments. In the second step, we specify the residual-augmented model of $\mathrm{E}[H_{it} | \mathbf{x}_{1it}, \mathbf{z}_{it}, \bar{y}_{it}, \nu_{it}] = \Phi\left(\varphi_k + \mathbf{x}_{1it}' \boldsymbol{\beta}_k + \bar{\mathbf{x}}_i' \boldsymbol{\theta}_k + \psi_k \ln \bar{y}_{it} + \rho_k \hat{\nu}_{it}\right)$. Here, too, the Chamberlain-Mundlak device concerns both the strictly exogenous variables and the excluded instruments. The subscript k denotes a new scale factor $(1 + \sigma_k^2)^{-1/2}$. The solution is to simply condition on an estimate of the omitted variable ($\hat{\nu}_{it}$). A test of $\hat{\rho}_k = 0$ corresponds to a test of exogeneity and does not depend on the first step under the null (see Hausman, 1978).

The asymptotic standard errors must be adjusted for the uncertainty of the first step and can be derived via the delta method or approximated with the panel bootstrap.

Accounting for unbalancedness adds another layer of complication. Contrary to linear models with CRE, where most estimators need little or no practical adjustments to work with unbalanced panels, estimates from non-linear CRE models are inconsistent if applied to unbalanced panel data. The main problem is that both the estimates and the variances of the correlated random effects can differ depending on the sample size. Wooldridge (2010a) proposes to also directly model this dependence. Assuming that selection is conditionally independent, we can extend our model by letting the unobserved effects vary depending on the panel size: $\mathrm{E}[\mu_i|\mathbf{w}_i] = \sum_{r=1}^{T} \delta_{T_i,r}\varphi_{kr} + \sum_{r=1}^{T} \delta_{T_i,r}\bar{\mathbf{x}}_i'\boldsymbol{\theta}_{kr}$, where $\mathbf{w}_i$ is a vector of functions of the conditioning variables sufficient to represent the distribution $\mathrm{D}[\mu_i|(s_{it}, s_{it}\mathbf{x}_{1it}, s_{it}\mathbf{z}_{it}, s_{it}\ln\bar{y}_{it})] = \mathrm{D}[\mu_i|\mathbf{w}_i]$; further, s_{it} is a selection indicator and $\delta_{T_i,r}$ is the Kronecker delta which is equal to unity if $T_i = r$ and zero otherwise. The coefficients are still scaled by $(1+\sigma_k^2)^{-1/2}$. Without further assumptions, this implies that we cannot use the observations where $T_i = 1$ as these have no separately identifiable panel dimension. Hence, they drop out of the estimating equation. Additionally, we also let the conditional variance depend on the panel size such that $\mathrm{var}(\mu_i|\mathbf{w}_i) = \sigma_\mu^2 \exp(\sum_{r=2}^{T-1} \delta_{T_i,r}\omega_r)$, where the omegas represent unknown variance parameters and σ_μ^2 is the variance of the unobserved heterogeneity when $T_i = T$.[10,11] The result is a variable scale factor: $(1+\sigma_\mu^2 \exp(\sum_{r=2}^{T-1} \delta_{T_i,r}\omega_r))^{-1/2}$.

A convenient reparameterization arises when we treat the overall variance as heteroskedastic and assume that $\mathrm{D}[\mu_i|\mathbf{w}_i]$ is normal (see Wooldridge, 2010a). Dividing the conditional expectation by $\exp(\sum_{r=2}^{T-1} \delta_{T_i,r}\tilde{\omega}_r)$, where $\tilde{\omega}_r$ denotes a new set of unknown parameters for the *overall* variance, we again obtain a constant scale factor. Then, the reparametricized two-step unbalanced CRE model is

$$\mathrm{E}[H_{it}|\mathbf{x}_i, \nu_{it}, \mathbf{w}_i] = \Phi\left(\frac{\mathbf{x}_{1it}'\boldsymbol{\beta}_h + \psi_h \ln\bar{y}_{it} + \rho_h\hat{\nu}_{it} + \sum_{r=2}^{T} \delta_{T_i,r}\varphi_{hr} + \sum_{r=2}^{T} \delta_{T_i,r}\bar{\mathbf{x}}_i'\boldsymbol{\theta}_{hr}}{\exp\left(\sum_{r=2}^{T-1} \delta_{T_i,r}\tilde{\omega}_r\right)^{1/2}}\right) \tag{13}$$

where the explanatory variables at t are $(1, \mathbf{x}_{1it}', \ln\bar{y}_{it}, \hat{\nu}_{it}, \delta_{T_i,2}\bar{\mathbf{x}}_i', \ldots, \delta_{T_i,T}\bar{\mathbf{x}}_i')$ and the variance depends on a set of dummy variables shifting from $T_i = 2$ to $T_i = T-1$ with $T_i = T$ as the base. The subscript h denotes the new scale factor. The specification nests the balanced case. If the panel is balanced, the numerator has only one set of time averages and a constant in addition to the time-varying covariates, while the denominator is unity. The first estimation step is also augmented to accommodate the varying panel sizes. We obtain the residuals via $\hat{\nu}_{it} = \ln\bar{y}_{it} - \mathbf{z}_{it}'\hat{\boldsymbol{\gamma}}_1 - \mathbf{x}_{1it}'\hat{\boldsymbol{\beta}}_1 - \sum_{r=2}^{T} \delta_{T_i,r}\hat{\varphi}_{1r} - \sum_{r=2}^{T} \delta_{T_i,r}\bar{\mathbf{x}}_i'\hat{\boldsymbol{\theta}}_{1r}$, where $\mathbf{x}_{1it}'$ includes time dummies and $\bar{\mathbf{x}}_i'$ their time averages. The heterogeneity related to the instruments (the $\bar{\mathbf{z}}_i$ time averages in $\bar{\mathbf{x}}_i$) is interacted with the panel size dummies and thus enters the first step flexibly. This, too, simplifies back to the earlier result in the balanced case if we remove the redundant variables (i.e. the averages of time effects).

[10] As Wooldridge (2010a) points out, it is possible to model the conditional expectation and variance even more flexibly by allowing for additional intercepts, trends, and variances/covariances to approximate the non-parametric relationship from Altonji and Matzkin (2005).

[11] We could also let the conditional variance depend on inequality (which can be motivated by assuming log normality of income). This relaxes an implicit assumption, namely that the marginal proportional rate of substitution ($MPRS_t = -\hat{\varepsilon}_t^{H\bar{y}}/\hat{\varepsilon}_t^{HG}$) is constant. The models fit marginally better but the substantive implications change very little. Additional results are available on request.

Obviously this is a complicated model to fit but it can be estimated by any software that has a heteroskedastic probit implementation without any restrictions on the dependent variable (Wooldridge, 2010a).[12] Since this is a quasi-maximum likelihood estimator (QMLE), the standard errors based on the inverse information matrix will be too conservative and need to be adjusted for clustering at the country level (for details see Papke and Wooldridge, 1996; 2008; Wooldridge, 2010b). Apart from the assumptions made to restrict the unobserved heterogeneity and endogeneity, fractional probit only requires correct specification of the conditional mean irrespective of the true distribution of the dependent variable (Gourieroux, Monfort, and Trognon, 1984). Hence, it is as robust as non-linear least squares but potentially more efficient.

We still need to define the average partial effects (APEs) and elasticities. Both can be derived from the average structural function (ASF) computed over the selected sample (see Blundell and Powell, 2004; Wooldridge, 2010a) which makes clear that *only the APEs of time-varying covariates are in fact identified.* Let the linear predictors inside the cumulative normal be $\mathbf{m}'_{it1}\hat{\boldsymbol{\xi}}_1$ for the main equation and $\mathbf{m}'_{it2}\hat{\boldsymbol{\xi}}_2$ for the variance equation. Then, $\widehat{\text{ASF}}(\mathbf{x}_t) = N^{-1}\sum_{i=1}^{N} \Phi\left(\mathbf{m}'_{it1}\hat{\boldsymbol{\xi}}_1/\exp(\mathbf{m}'_{it2}\hat{\boldsymbol{\xi}}_2)^{1/2}\right)$, where $\mathbf{x}_t$ refers to all time-varying covariates including mismeasured income, and the coefficients are the scaled QMLE estimates. We always need to average over the cross-section dimension in order to get rid of the unobserved effects, varying panel sizes, and endogeneity/ measurement error. The APE at time t of a particular continuous variable is simply the derivative of the ASF with respect to that variable. We usually plug in interesting values for $\mathbf{x}_t$ and obtain the APEs assuming the entire sample shares these characteristics.

Analogously, the elasticity with respect to any $x_k \in \mathbf{x}_t$ (provided that x_k is in logs and does not show up in the variance equation) is

$$\hat{\varepsilon}_t^{Hx_k} = \hat{\beta}_k \times N^{-1}\sum_{i=1}^{N} \exp\left(-\mathbf{m}'_{it2}\hat{\boldsymbol{\xi}}_2/2\right)\lambda\left(\mathbf{m}'_{it1}\hat{\boldsymbol{\xi}}_1/\exp(\mathbf{m}'_{it2}\hat{\boldsymbol{\xi}}_2)^{1/2}\right) \tag{14}$$

and the semi-elasticity ($\hat{\eta}_t^{Hx_k}$) is the derivative of the ASF with respect to x_k; that is, the average partial effect (APE). We also average over time in order to obtain a scale factor.

The basic structure is exactly the same as in the simpler versions derived in the previous section with the addition of a variance equation adjusting for the degree of unbalancedness. If the panel is balanced, the non-redundant sums inside the linear predictors simplify and we again obtain the CRE analogues of eqs. (7) and (8).

3 Data

Based on the World Bank's PovcalNet database[13], we compile a new and comprehensive data set consisting of 809 nationally-representative surveys spanning 124 countries from 1981 to 2010.[14] Smaller panels of this data have been used in previous studies (e.g. Chambers and Dhongde, 2011; Kalwij and Verschoor, 2007; Adams, 2004) and the World Bank's methodology is described in more detail in Chen and Ravallion (2010). Here we

[12]However, most implementations (e.g. Stata's `hetprob`) only allow binary dependent variables. We implement the estimator in a new module called `fhetprob` (see Bluhm, 2015, *forthcoming*).

[13]The data is publicly available at `http://iresearch.worldbank.org/PovcalNet` (last accessed May 20 2013, last updated April 18 2013).

[14]Supporting materials and the panel data set are available at `www.richard-bluhm.com/data/`.

only briefly summarize the main features.

All data originate from household surveys. Our primary measure of poverty is the headcount index (H_{it}) given a poverty line (z) of 2$ a day or 60.83$ a month. The poverty headcount at the 1.25$ a day or 38$ a month poverty line is used as a secondary measure. The data also contains measures of mean monthly per capita household income or consumption expenditures ($\bar{y}_{it}$)[15], the Gini coefficient of inequality (G_{it}) for income or consumption expenditures, the surveyed population (pop_{it}),[16] and a set of indicators distinguishing if the survey uses income or consumption as a welfare measure and whether unit-level or grouped-level (deciles or finer quantiles) data are used. About 63% of the data come from expenditure surveys and about 74% are estimated from grouped data. All monetary quantities are in constant international dollars at 2005 PPP-adjusted prices.

Some countries[17] do not conduct nationally representative surveys but instead report urban and rural data separately. We simply weigh the poverty and income data using the relative urban/rural population shares to construct national series. Since the Gini is not subgroup decomposable, we employ a mixing of two log normal distributions approximation to estimate a national Gini coefficient (Young, 2011).[18] If only one urban or rural survey is available in any given year, we usually drop the survey, except in the case of Argentina where urbanization is near or above 90% for most of the sampled period and we thus consider the urban series nationally representative. This results in an unbalanced panel of 124 countries, with an average time series ($\bar{T}$) of about 6.5 surveys for a total of 809 observations, spanning 30 years. Table 1 provides summary statistics for the entire panel. In Appendix B, Table B-1 presents summary statistics by region, and List B-1 lists the countries and the corresponding numbers of surveys in the sample.

Table 1 – Summary statistics

	Mean	Std. Deviation	Min	Max	N
Main variables					
H_{it} – Headcount (2$)	0.303	0.286	.0002	.9845	809
H_{it} – Headcount (1.25$)	0.182	0.219	.0002	.9255	789
G_{it} – Gini coefficient	0.424	0.102	.2096	.7433	809
$\bar{y}_{it}$ – Mean income or expenditure in $ per month	194.59	125.90	14.93	766.78	809
PCE^P_{it} – Consumption (PWT) in $ per month	338.64	234.59	14.39	1231.21	795
Survey type dummies					
Consumption (Grouped)	0.611	0.488	0	1	809
Consumption (Unit)	0.015	0.121	0	1	809
Income (Grouped)	0.132	0.339	0	1	809
Income (Unit)	0.242	0.429	0	1	809

For the linear models in log differences, we also construct a second data set of 'poverty spells' at the country level (data in changes). Manually defining poverty spells serves two purposes. First, we only use data from runs of the same survey type to avoid introducing

[15] Computed as a simple per capita average without equivalence scaling.

[16] Several entries in the PovcalNet population data are clearly mistaken, zero or missing. We fix the values/series using data from the World Development Indicators.

[17] China, India and Indonesia.

[18] PovcalNet omits weighting some recent data. To use a single consistent method, we apply Young's formula in all cases where separate urban and rural surveys are combined. The approximation is very accurate. The formula is $G = \sum_{i=1}^{K}\sum_{j=1}^{K} \frac{s_i s_j \bar{y}_i}{\bar{Y}} \left(2K\left[\frac{\ln \bar{y}_i - \ln \bar{y}_j + \frac{1}{2}\sigma_i^2 + \frac{1}{2}\sigma_j^2}{(\sigma_i^2+\sigma_j^2)^{1/2}}\right] - 1\right)$ where K is the total number of subgroups, s_i is the population share of the i-th subgroup and $\bar{Y}$ is the population-weighted mean income across all subgroups.

artificial changes when there is a switch from income to expenditure surveys or vice versa (same for unit and grouped-level data; we assume the grouping does not change). Second, we annualize all differences to mitigate any biases arising from estimating elasticities over very different time periods. This leads to a smaller data set of 648 observations in 104 countries, as differencing requires $T_i \geq 2$. Most studies using poverty differences lose additional observations when the poverty headcount is zero at the beginning or end of a spell (e.g. Kalwij and Verschoor, 2007), implying a relative change in the poverty headcount of negative or positive infinity. This problem does not occur in our data, but more importantly, our approach does *not* require such ad hoc adjustments.

We add per capita consumption data from national accounts to the survey-based panel, which we later use as instruments for mean survey income and require for the poverty projections. Personal consumption expenditures are retrieved from both from the World Development Indicators (WDI) and Penn World Table 7.1 (PWT), denoted PCE_{it}^{W} and PCE_{it}^{P} respectively.[19] The PWT version is preferred in the estimations; for the poverty projections a "merged" series (PCE_{it}) is constructed using the WDI series as a benchmark but replacing it with PWT data if coverage over 1981-2010 is better. Both series are in constant 2005 prices, but the PWT adjust the original 2005 ICP data and interpolate differently between benchmark years (?). For the projections, we also use estimates of the total population from 2010 to 2030 based on the medium fertility variant from the World Population Prospects (the 2010 revision).

4 Results

4.1 Fractional response models

Table 2 presents our main results, with each specification progressively addressing more estimation issues (unobserved effects, unbalancedness and measurement error). In all specifications, we include time averages to proxy for time invariant measurement differences across different countries (unobserved effects). We also include survey type dummies (consumption or income, grouped or unit data) as reported poverty is typically lower in income surveys than consumption surveys and the availability of grouped versus unit-level data in PovcalNet may proxy for other systematic survey differences. In addition, time dummies allow for unspecified common year effects.

Model (1) includes correlated random effects but entirely ignores unbalancedness. As expected, the coefficient on average income is negative and the coefficient on inequality is positive. Since the estimated coefficients are arbitrarily scaled, the adjacent column reports the average partial effects (APEs) and we report the scale factor separately in the bottom panel. The APEs in model (1) imply the following semi-elasticities. On average, one *percent* income growth leads to a 0.284 *percentage point* reduction in the population that is poor and a corresponding one *percent* increase in inequality leads to a 0.232 *percentage point* increase in poverty. Turning to elasticity concepts, the average income elasticity across the entire estimation sample is about -1.83 (SE = 0.084) and the average Gini elasticity is about 1.5 (SE = 0.167). Hence, a one *percent* increase in income leads to about a 1.83 *percent* decrease in poverty. These two estimates are located

[19] Monthly PCE_{it}^{P} is computed as $(kc_{it}/100 \times rgdpl_{it}/12)$, where kc_{it} is the consumption share and $rgdpl_{it}$ is GDP per capita (Laspeyres) in 2005 constant prices. Similarly, PCE_{it}^{W} is household final consumption expenditure in 2005 prices divided by the population and converted to monthly figures.

near the lower bound of the results typically found in the literature.[20]

Our first specification could be biased due to the strong unbalancedness of the panel and the presence of time-varying measurement error in income and inequality. Model (2) addresses unbalancedness by including panel size dummies, interactions of the time averages with the panel size dummies, and a separate variance equation. We consider this our best specification *without* correcting for measurement error. The substantive conclusions change very little. The APE of income is virtually unchanged and the APE of inequality increases by about one standard error. Clearly, varying panel sizes introduce little bias on average. Nonetheless, they may still have a non-negligible effect on the (semi-)elasticites at particular points in time.

Table 2 – Fractional probit models (QMLE) – Dependent variable: H_{it}, 2$ a day

	(1) Regular		(2) Unbalanced		(3) Unbalanced + Two-Step	
	H_{it}	APEs	H_{it}	APEs	H_{it}	APEs
$\ln \bar{y}_{it}$	-1.263	-0.284	-0.880	-0.281	-1.049	-0.339
	(0.054)	(0.012)	(0.048)	(0.011)	(0.198)	(0.035)
$\ln G_{it}$	1.032	0.232	0.786	0.251	0.775	0.251
	(0.114)	(0.026)	(0.098)	(0.026)	(0.163)	(0.032)
$\hat{\nu}_{it}$					0.133	
					(0.113)	
CRE (Corr. Rand. Effects)	Yes		Yes		Yes	
Survey type dummies	Yes		Yes		Yes	
Time dummies	Yes		Yes		Yes	
Panel size dummies	No		Yes		Yes	
Panel size dummies × CRE	No		Yes		Yes	
Variance equation	No		Yes		Yes	
Scale factor	0.225		0.319		0.323	
$N \times \bar{T}$	789		789		775	
N	104		104		103	
pseudo R^2	0.992		0.996		0.997	
$\ln \mathcal{L}$	-219.3		-315.6		-313.4	
$\sqrt{MSE}$	0.0355		0.0238		0.0235	

Notes: 20 observations with $T_i = 1$ not used in estimation. In models (1) and (2), the standard errors of the coefficients are robust to clustering at the country level and the standard errors of the APEs are computed via the delta method. We include the time averages of the survey type and time dummies in (2) and (3), but constrain their coefficients to be equal across the panel sizes. The standard errors of the coefficients and the APEs in model (3) account for the first stage estimation step with a panel bootstrap using 999 bootstrap replications. The linear projection in the first stage uses $\ln PCE^P_{it}$ as an instrument for $\ln \bar{y}_{it}$. The first-stage cluster-robust F-statistic in (3) is 28.05. Model (3) also excludes West Bank and Gaza entirely (2 observations) and 12 observations from ECA countries pre-1990 for lack of PCE data.

Our preferred specification, model (3), is the empirical counterpart of the two-step estimator presented in eq. (13). We account for measurement error in income by

[20]The typical range for the income elasticity in earlier studies is from about 2 to 5, while the range for the Gini elasticity is much wider. Newer studies suggest the income elasticity is closer to 2. This is largely owed to parameter instability in linear approximations and changing data coverage. However, as Chambers and Dhongde (2011) report, estimates of the income elasticity using the new 2005 PPPs are also universally lower (in absolute value) than estimates based on the earlier 1993 PPPs.

instrumenting survey mean incomes or expenditures with per capita consumption from the national accounts (PCE_{it}^P). The main identifying assumption is that any measurement error in per capita consumption from the national accounts is orthogonal to survey-based measurement error in income or expenditures. As both expenditure measures are estimated very differently in practice, this is a plausible and common identification strategy (Ravallion, 2001). Figure B-3 in Appendix B highlights the strength of the first stage relationship. It shows a partial regression plot of mean incomes or expenditures from the surveys against per capita consumption from the National Accounts, after taking out the variation in the Gini, the time averages of the Gini and PCE_{it}^P, panel size dummies, survey dummies and time dummies.

Ignoring the first-stage variability of income, we find some evidence of measurement error (cluster robust t-stat $\approx$ 1.69).[21] After accounting for the two-step nature the evidence for measurement error is considerably weaker (panel bootstrap t-stat $\approx$ 1.18). The APE of income, on the other hand, is larger in absolute value, hence we conclude that the coefficient of income in models (1) and (2) is attenuated towards zero. These results suggest that classical attenuation bias is more of a problem than systematic survey bias due to under-reporting or over-reporting of incomes and expenditures. Since the biases arising from these two types of measurement error are likely to run in opposing directions, our estimates imply that they are not fully offsetting as in Ravallion and Chen (1997).

The unbiased average-income semi-elasticity of poverty estimated in model (3) is considerably larger in absolute value than in the previous two specifications. A one percent increase in incomes leads to a 0.339 percentage point reduction in the population that is poor. Likewise, the average elasticity increased ($\bar{\varepsilon}^{H\bar{y}} \approx 2.21$, SE = 0.156) and is now closer to the conventional estimates of about two. The inequality elasticity remains about the same ($\bar{\varepsilon}^{HG} \approx 1.64$, SE = 0.188). In fact, there may also be non-negligible measurement error in observed inequality as measured by the Gini coefficient. However, since inequality is most often estimated from household surveys and estimates based on alternative sources such as tax records are not available on a cross-country basis, we are lacking a corresponding instrument for the Gini coefficient.

Is there evidence of additional non-linearity missing in the non-linear functional form of the poverty headcount? To examine this possibility, we add squares $(\mathbf{m}_{it1}'\hat{\boldsymbol{\xi}}_1)^2$ and cubes $(\mathbf{m}_{it1}'\hat{\boldsymbol{\xi}}_1)^3$ of the linear predictors in models (1) and (2) for a RESET-type test as suggested by Papke and Wooldridge (1996). In the most basic model (1), this yields a robust χ_2^2-statistic of 4.65 for a p-value of 0.098, giving no reason for concern. For the heteroskedastic model (2), the RESET test provides some evidence of additional non-linearity (robust $\chi_2^2 = 9.15$, $p > 0.01$). However, there is no theoretical reason to expect additional powers or interactions of income and inequality to enter the model.

At first sight, all three models may suggest an only moderately larger effect of income growth when holding inequality constant relative to changes in inequality, and *vice versa*. This does not imply that both variables have the same scope for change or have to change independently for that matter. We estimate the impact of each component and not its contribution to overall poverty reduction (for estimates of the contributions see Kraay, 2006). While there is substantial variation in inequality, it shows no systematic trend over the sample period from 1981 to 2010.[22] In contrast, incomes and expenditures have increased substantially in all regions over the same time span (see Table B-5 in

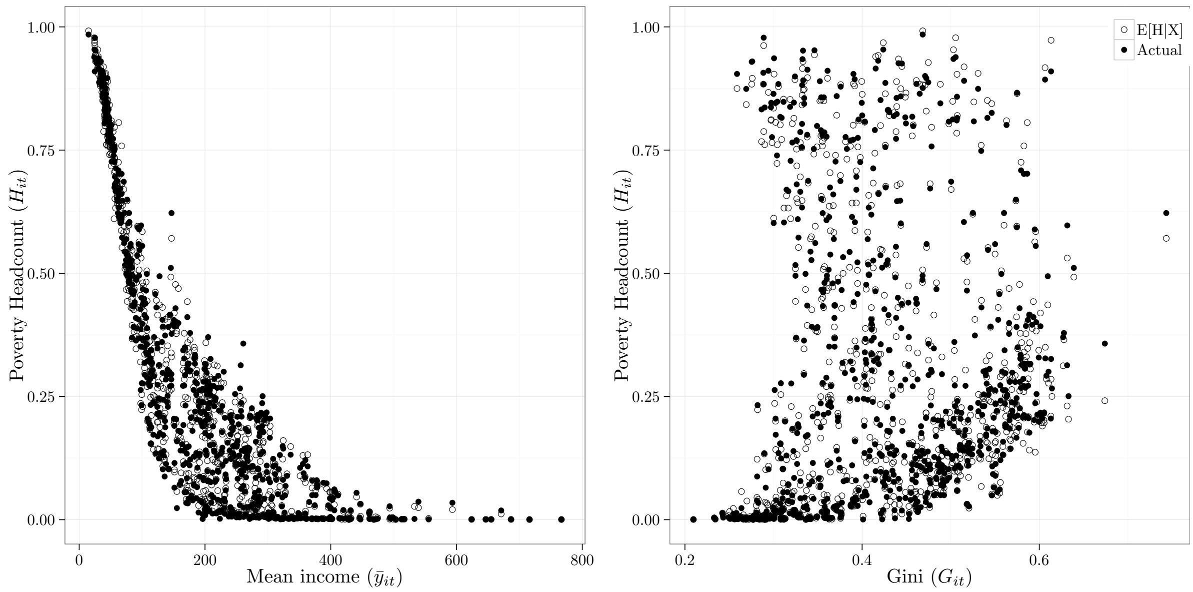
[21]As mentioned earlier, the Hausman (1978) test does not depend on the first stage under the null.

[22]In a simple regression of the Gini coefficient on time, we cannot reject the null hypothesis that the time trend is zero (cluster robust t-stat $\approx$ 0.07 and $p > 0.94$).

Appendix B). Moreover, the effect of income growth is not constant. In these models, it depends strongly on the levels of inequality and income. Thus there is a 'double dividend' to improvements in distribution (Bourguignon, 2003) and substantial heterogeneity in the estimated poverty (semi-)elasticities across time and space – an issue to which we return shortly.

Perhaps the most striking fact about all three specifications is how well they fit. For more intuitive comparisons, the last row shows the square root of the mean squared residual for each model – a model metric suggested by Wooldridge (2010a). Already in the first model, we predict the observed poverty headcount for each country-year with about three and a half percentage points accuracy and with better than two and a half percentage points accuracy in the preferred specification. This truly reflects an identity relationship. A simple pseudo-R^2 measure of the correlation between the observed and fitted values for models (1) to (3) suggests near perfect fit ($R^2 > 0.99$). Figure 1 illustrates this point and shows the shape of the estimated effects. Using our preferred specification, we plot both the observed headcount and the predicted headcount over the range of observed mean income or expenditures (left panel) and inequality (right panel). The quality of the non-linear approach is readily apparent as the fit is very close at either bound (near unity or near zero) and the model does not predict nonsensical values. Further, the variation in the observed values is completely covered by the model predictions. In linear models, neither of these two outcomes is guaranteed.

Figure 1 – Data versus fitted values, preferred specification, 2$ a day



For comparison, Table B-2 in Appendix B reproduces the linear approach of the previous literature using the data in levels and the poverty spell data in differences. The differences in the estimated *average* elasticities are not large, as is typical for comparisons between linear and non-linear approaches. We do not discuss these results in detail since they suffer from the expected problems (see Section 2). First, when switching from fixed effects to annualized differences in the simple models with only income and inequality, measurement error increases and attenuates the income coefficient. Second, the models with interaction terms do not fit nearly as well as those reported earlier and many coefficients are insignificant. Third, the two-step GMM results for the interaction models are unstable and not able to convincingly solve the problem of measurement error. The last model, which is in the spirit of the preferred specification of Kalwij and Verschoor (2007), even implies a negative Gini elasticity and all coefficients are estimated with great imprecision. In sum, these models do not perform well in comparison to their fractional response counterparts and are thus unlikely to produce reliable estimates over a wide range of circumstances.

Conversely, the strength of the fractional response approach lies in its ability to deliver much more precise estimates of effects other than the overall mean. Table 3 and Table 4 illustrate this point by estimating the income elasticities and Gini elasticities over different time periods for the six geographic regions in our sample. They are computed according to eq. (14) by plugging in the different time period and region specific means of income ($\ln \bar{y}_{it}$) and inequality ($\ln G_{it}$), and then averaging over the entire sample population. The standard errors of the elasticities are computed via a panel bootstrap and thus account for the uncertainty of the first stage. We present regional and temporal elasticities here but also provide estimates for the semi-elasticities in Table B-3 and Table B-4 in Appendix B for comparison.

Table 3 – Predicted regional income elasticities, preferred specification, 2$ a day

	Time period				
	1981–1989	1990–1994	1995–1999	2000–2004	2005–2010
East Asia and Pacific	-0.991	-1.029	-1.237	-1.139	-1.578
	(0.030)	(0.033)	(0.055)	(0.043)	(0.101)
Eastern Europe and Central Asia	-4.358	-2.892	-2.700	-2.846	-3.304
	(0.555)	(0.309)	(0.277)	(0.304)	(0.384)
Latin America and Caribbean	-2.284	-2.374	-2.425	-2.349	-2.985
	(0.243)	(0.257)	(0.271)	(0.258)	(0.366)
Middle East and North Africa	-2.176	-2.116	-2.024	-1.966	-2.501
	(0.203)	(0.188)	(0.168)	(0.161)	(0.246)
South Asia	-0.548	-0.629	-0.810	-1.024	-1.192
	(0.053)	(0.048)	(0.030)	(0.032)	(0.046)
Sub-Saharan Africa	-0.831	-0.437	-0.436	-0.592	-0.632
	(0.027)	(0.039)	(0.040)	(0.035)	(0.033)

Notes: Standard errors obtained via a panel bootstrap using 999 replications. The predicted elasticities are based on estimated APEs at each region/time-period mean of $\ln \bar{y}_{it}$ and $\ln G_{it}$.

There is considerable regional and temporal heterogeneity in the estimated income elasticities. However, its origins are very mechanical. As the theoretical derivations in Section 2 show and our estimates make clear, the income elasticity is an increasing function of income. In other words, regional heterogeneity of the income elasticity is mainly due to income heterogeneity. More affluent regions (Eastern Europe and Central Asia, Latin America and the Caribbean, and the Middle East and North Africa) have higher income elasticities than poorer regions (East Asia and Pacific, South Asia and Sub-Saharan Africa). Income dynamics over time are also clearly visible. In Eastern Europe and Central Asia, for example, income is comparatively high before the post-communist transition, sharply collapses throughout the 1990s and then recovers during the 2000s. Compared to earlier results (e.g. Kalwij and Verschoor, 2007), we find markedly higher average income elasticities in more affluent regions and lower elasticities in poorer regions. Throughout Table 3, the standard errors are small compared to the point estimates and *remain* very accurate for regions with more extreme values (e.g. very low income and above average inequality in Sub-Saharan Africa in the 1980s).

Table B-3 in Appendix B presents the region and time specific income semi-elasticities of poverty. There the picture is reversed. Comparatively more affluent regions have less people near the poverty line, and thus the poverty reduction potential from a one percent increase in incomes is much smaller in terms of the population lifted out of poverty. This dynamic is again best visible in Eastern Europe and Central Asia, where absolute poverty at the 2$ a day poverty line is almost non-existent just before the post-communist transition and then rises sharply in the 1990s as incomes decline. Correspondingly, the semi-elasticity is close to zero in the 1980s but then it increases as more people fall into poverty. Likewise, the biggest poverty reduction potential in 2005-2010 was in East Asia, South Asia and Sub-Saharan Africa. This highlights an important point. For development policy, we really care more about poverty reduction in terms of the percent of the population lifted out of poverty rather than relative changes in the poverty headcount. Hence, semi-elasticities are the pertinent concept (see also Klasen and Misselhorn, 2008).

Table 4 – Predicted regional Gini elasticities, preferred specification, 2$ a day

	Time period				
	1981–1989	1990–1994	1995–1999	2000–2004	2005–2010
East Asia and Pacific	0.732	0.760	0.914	0.841	1.165
	(0.105)	(0.101)	(0.113)	(0.108)	(0.144)
Eastern Europe and Central Asia	3.219	2.136	1.994	2.102	2.440
	(0.510)	(0.307)	(0.283)	(0.296)	(0.353)
Latin America and Caribbean	1.687	1.753	1.791	1.735	2.205
	(0.186)	(0.198)	(0.199)	(0.189)	(0.269)
Middle East and North Africa	1.607	1.563	1.495	1.452	1.847
	(0.197)	(0.198)	(0.196)	(0.185)	(0.253)
South Asia	0.405	0.464	0.598	0.756	0.880
	(0.093)	(0.097)	(0.095)	(0.107)	(0.127)
Sub-Saharan Africa	0.614	0.322	0.322	0.437	0.467
	(0.087)	(0.055)	(0.060)	(0.066)	(0.069)

Notes: Standard errors obtained via a panel bootstrap using 999 replications. The predicted elasticities are based on estimated APEs at each region/time-period mean of $\ln \bar{y}_{it}$ and $\ln G_{it}$.

The region and time specific Gini elasticities in Table 4 show where the potential of redistributive policies in terms of proportionate reductions in the poverty headcount was largest over the last three decades. Unsurprisingly, these regions are Eastern Europe and Central Asia, Latin America and the Caribbean, and the Middle East and North Africa – all of which have above average inequality. Sub-Saharan Africa starts out with high inequality in the 1980s[23] but incomes are very low relative to the poverty line, so that the Gini elasticity is small. This is the flip side of the dependency on initial levels: countries can be so poor and unequal that the immediate effects of equalization and income growth on *relative* changes in the poverty headcount are comparatively small. However, here too, the semi-elasticities presented in Table B-4 in the Appendix help to clarify the picture. There the relative position of poorer and richer countries is reversed. The potential for reducing poverty through redistribution in terms of percent of the population that is poor

[23] The population-weighted mean Gini in the 1980s is 0.4608.

was larger in poorer regions throughout the entire period from 1981 to 2010.

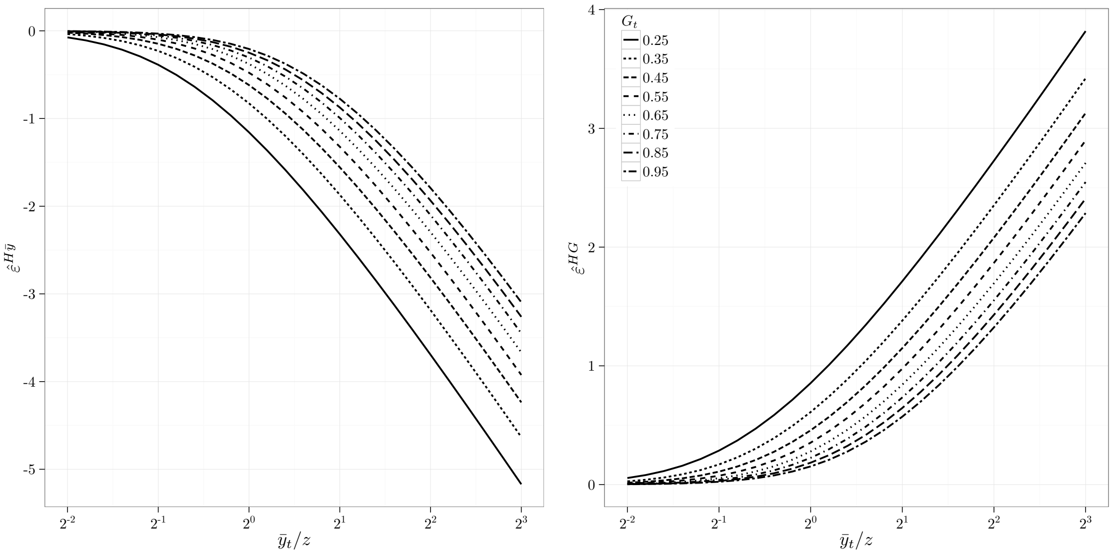
The 'double dividend' of reductions in inequality is illustrated in Figure 2 by graphing the estimated poverty elasticities or semi-elasticities over different combinations of income and inequality. Again, we compute these estimates according to eq. (14) by plugging in the different values for mean income or expenditures ($\ln \bar{y}_{it}$) and inequality ($\ln G_{it}$), and then averaging over the entire sample. As Figure 2a illustrates, on top of the direct poverty alleviating effect of income redistribution towards the poor, a lower level of inequality also raises the income elasticity in absolute value at every point. However, the magnitude of both elasticities is steeply increasing in the level of income; that is, the return to either income growth or equalization is bigger, the higher the income level. This may invite the conclusion that growth matters more at lower levels of income, while redistribution is only important for high income and high inequality countries. This, precisely, is the misleading feature of relative changes.

Figure 2b shows the predicted income and Gini semi-elasticities of poverty. The picture is very different and in many ways more intuitive. If the shortfall is too large – the mass of the income distribution is too far to the left of the poverty line – then both the income and the Gini semi-elasticities approach zero. However, if the country is very affluent – the mass of the income distribution is far to the right of the poverty line – then both semi-elasticities also approach zero. In between those two extremes, improvements in the income distribution can make a very large difference in terms of percent of the population lifted out of poverty, both, directly through redistribution and indirectly through growth. At $\bar{y}_t/z = 1$, for example, a Gini of 0.25 implies that one percent income growth leads to a 0.584 percentage point reduction in the poverty headcount and a Gini of 0.55 implies that one percent income growth leads to a 0.378 percentage point reduction in the poverty headcount. Especially at very low average income levels the initial income distribution is decisive. It practically determines whether there is any substantial potential for poverty alleviation through income growth at all (in terms of percent of the population that is poor). Moreover, as the Gini semi-elasticity also depends on the level of inequality, further improvements in the income distribution will have a larger effect on poverty reduction at lower levels of inequality. Contrary to Figure 2a, this suggests that poverty reduction strategies should focus both on income growth *and* equalization, especially in least developed countries and high inequality countries where the total returns to redistribution are large. Again, for policy questions, these relationships are much more pertinent than relative changes in the poverty headcount.

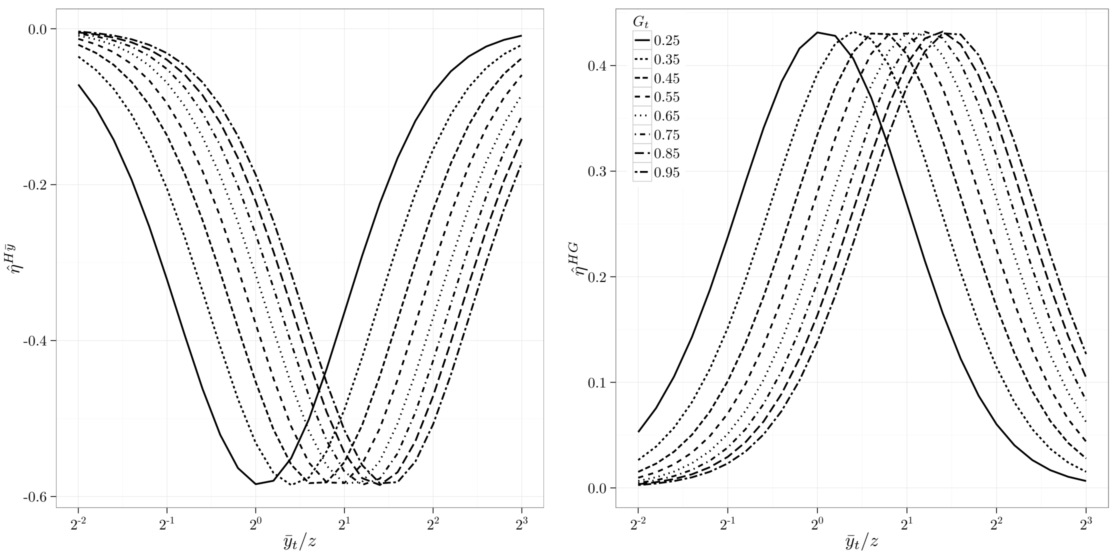
Could the decomposition be improved by allowing for other "more ultimate" determinants of poverty? If the assumption of log normality is justified, mean income and the Gini fully describe the distribution of incomes and expenditures, and there is logically no scope for other variables to enter. Yet this assumption is restrictive and we deliberately do not rely on log normality. In fact, we expect it to be violated at least for some cases (see, e.g., the host of alternative distributions analyzed by Bresson, 2009). More realistic distributions usually have more than one shape parameter to better capture skewness, long tails or the existence of multiple modes. "Ultimate factors" could thus be proxies for systematic deviations from equiproportional shifts in the distribution of incomes and expenditures. Weak institutions, for example, may explain the fact that the rich receive more of the gains. Table B-6 in Appendix B extends the heteroskedastic fractional probit models with data on institutions, human capital, access to credit and trade openness. The APEs of income and inequality are not affected by the inclusion of additional covariates and the APEs of other determinants are virtually zero. Thus we

Figure 2 – Predicted income and Gini elasticities and semi-elasticities of poverty, 2$ a day

(a) elasticities ($\hat{\varepsilon}^{H\bar{y}}$ and $\hat{\varepsilon}^{HG}$)



(b) semi-elasticities ($\hat{\eta}^{H\bar{y}}$ and $\hat{\eta}^{HG}$)



conclude that with only two variables, several dummies and correlated random effects, these specifications are essentially saturated. The fractional response approach leaves little room for misspecification of the decomposition.

While the literature on poverty reduction has produced mixed results so far, it is largely consistent with this view. Prominent examples are two studies by Dollar and Kraay (2002, 2004), who find that trade, inflation and other factors influence the incomes of the poorest quintile, while several other variables do not. However, they emphasize that these effects run predominantly through growth of GDP per capita. The interesting link is between some factor X and income or inequality, not between X and a measure of poverty. Usually such a “first stage” relationship has a dedicated literature that explicitly attempts to resolve causality issues and provides an appropriate theoretical background. Thus if we are interested in the effects of, say, institutions on poverty it is not only sufficient but, in our opinion, much more relevant to investigate the effects of institutions on income (e.g. see Acemoglu et al., 2001; Acemoglu and Johnson, 2005) and inequality (e.g. see Engerman and Sokoloff, 1997; Easterly, 2007). Separating these two estimation steps is important, as the impacts of income and distributional changes themselves depend on the initial levels of income and inequality.

4.2 Projecting poverty

Parts of the previous literature highlight that estimates of income and inequality elasticities or semi-elasticities are particularly useful for poverty simulations (e.g. Klasen and Misselhorn, 2008) and hence model fit is very important. Fractional response models provide a new, powerful and simple method of predicting poverty.

To illustrate the appeal and accuracy of this approach, we compare the predictions of our model for 2010 to the official World Bank data and then extrapolate poverty well into the medium-term future until 2030. Clearly, this is a hypothetical exercise and is not intended to replace any official estimates by the World Bank or national authorities. Rather it allows us to make somewhat more sophisticated predictions than back-of-the-envelope trend extrapolations and can provide a useful benchmark for setting global poverty reduction goals. Using fractional response models for this purpose has the added advantage that we can predict poverty responses to any combination of shifts in mean income and inequality. Further, these models have the attractive feature that the implied changes in the elasticities of poverty at different income and inequality levels are automatically taken into account.

In fact, even the official World Bank regional poverty figures involve a considerable amount of interpolation and extrapolation since most household surveys are not undertaken annually (for details see Chen and Ravallion, 2004). The basic steps are as follows. The World Bank first calculates poverty in the given survey year by fitting Lorenz curves to either the unit-level or grouped-level data. Then, average real household income is lined up to a reference year by interpolating between surveys or extrapolating with the growth rate of personal consumption expenditures per capita (PCE_{it}). Afterwards, the poverty headcount is recalculated using the new income level and the same Lorenz curve as before. If two surveys are available, one before and one after the reference year, the poverty headcount is a weighted average of the two estimates for the reference year.

Our method is similar in spirit. We proceed in four steps. First, we extrapolate the last available survey income to 2010 using actually observed country growth rates in PCE from the WDI, or PWT if the former is missing. Inequality is kept constant at the latest observed value. Second, we project mean income into the future using each country's average growth rate of PCE over the last 15 years. We assume that growth is distribution neutral, which is in line with the absence of any significant correlation between changes in inequality and income growth (see Figure B-4 in the Appendix). Third, we predict the poverty headcount in 5-year intervals from 2010 to 2030 using our preferred specification without adjusting for measurement error in income (Model 2 in Table 2).[24] We typically do not need estimates of the measurement errors in income or inequality for forecasting purposes, but we implicitly assume that their contribution remains stable over time. Finally, the world total and the regional aggregates are estimated as population-weighted

[24]Note that although our preferred specification is only estimated on the sub-sample where $T_i \geq 2$, we can use the model estimates to predict poverty for the entire sample ($T_i \geq 1$). We only lack estimates of the panel size effects for $T_i = 1$, so we assign these observations to the adjacent group ($T_i = 2$).

averages of our country level estimates using population data from the World Population Prospects. We do not provide standard errors for the point estimates since these are subject to fundamental uncertainty in the assumed PCE growth rates.

Table 5 shows the results and Table B-5 in Appendix B provides regional PCE growth rates highlighting the assumptions behind these forecasts. The comparison of the official World Bank poverty figures and our estimates in 2010 illustrates that our approach produces meaningful results. For three regions, our estimates are within a percentage point of the official figures, for another two regions, they are within less than 2.5 percentage points, and only for East Asia and Pacific, we estimate a much lower poverty level in 2010. Our results closely match the World Bank's results for the world total. Using the World Bank population data, our estimates imply 2,383.43 million people under the 2$ line worldwide in 2010 versus 2,395.21 million as reported by the World Bank.

Table 5 – Poverty projections ($\hat{H}_{it} \times 100$), preferred specification, 2$ a day

	2010 Official	2010 Estimate	2015 Estimate	2020 Estimate	2025 Estimate	2030 Estimate
East Asia & Pacific	29.73	26.70	16.80	10.53	6.92	4.86
Europe & Central Asia	2.35	2.98	1.88	1.17	0.73	0.46
Latin America & the Caribbean	10.37	10.62	8.96	7.59	6.46	5.52
Middle East & North Africa	12.04	14.57	11.36	8.86	6.92	5.41
South Asia	66.71	68.36	57.50	46.09	35.29	26.06
Sub-Saharan Africa	69.87	69.02	64.74	60.73	57.20	54.02
World Total	40.67	40.47	33.64	28.01	23.55	20.09

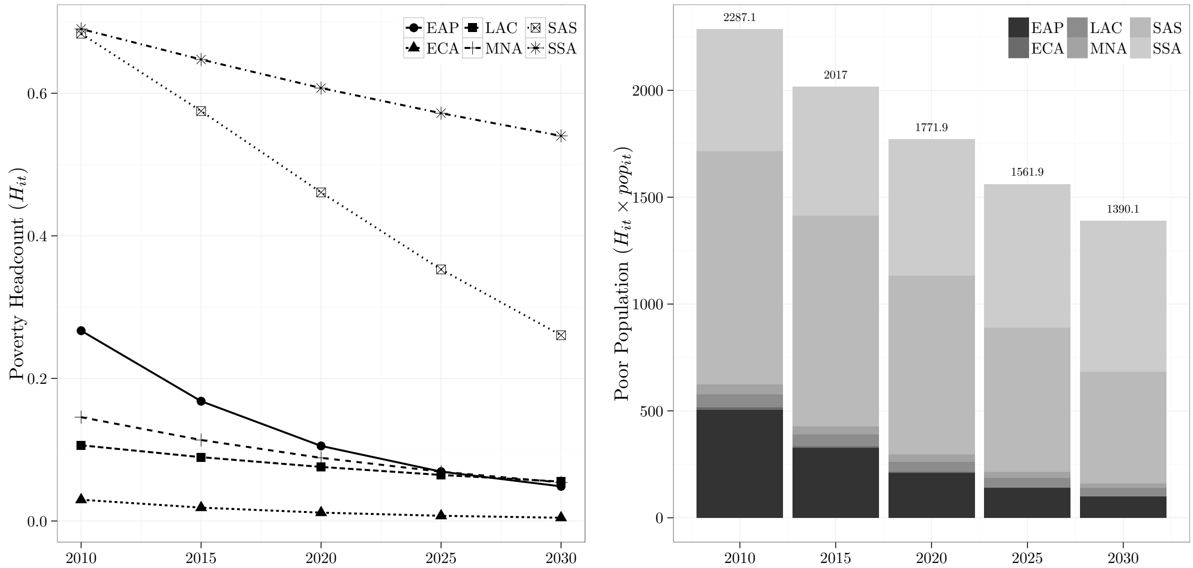
Notes: Regional aggregates are based on the World Bank classification of low and middle income countries in 1990. High income countries in 1990 are assumed to have no poor. The projections are for 123 countries. West Bank and Gaza is excluded as we lack both PCE and population data.

The anticipated regional trends from 2010 to 2030 are highlighted in Figure 3. The left panel shows the regional poverty rates and the right panel plots the regional distribution of the poor population. Given past growth trajectories, poverty in Sub-Saharan Africa and South Asia remains the fundamental development challenge of the twenty-first century. Estimated poverty in Sub-Saharan Africa is very high in 2010 (69.02%) and projected to remain high through 2030 (54.02%) on the entire subcontinent in spite of sustained income growth (about 2.3% p.a.). In South Asia, too, poverty is equally high in 2010 (68.36%) but projected to fall by more than half (to 26.06% in 2030). By 2030, about half of the world's poor will live in Sub-Saharan Africa, followed closely by South Asia.

Poverty in the East Asia and Pacific region, on the contrary, largely takes care of itself if incomes and consumption expenditures keep growing at the impressive rates of the last 15 years. We project poverty in East Asia (4.86%) to be below poverty levels in Latin America (5.52%) by 2030, and second only to Eastern Europe and Central Asia where absolute poverty virtually disappears (down to 0.46%). Most of the progress in East Asia is due to rapid income and expenditure growth in China. However, this prediction may not hold if a middle-income slow-down occurs in China as some observers suggest (see, e.g., Eichengreen, Park, and Shin, 2013).

Progress in Latin America and the Caribbean, and the Middle East and North Africa is noticeably slower in spite of the assumption of robust yet moderate income growth

Figure 3 – Predicted regional poverty, 2$ a day, 2010 to 2030



Notes: Regional aggregates are based on the World Bank classification of low and middle income countries in 1990. High income countries in 1990 are assumed to have no poor. The projections are for 123 countries. West Bank and Gaza is excluded as we lack both PCE and population data. Population data is the medium fertility variant from the World Population Prospects.

(about 2.2% and 2.9% p.a., respectively) and comparatively large income elasticities. This suggests that the countries in these regions should reinforce their poverty alleviation efforts. However, for the Middle East and North Africa these numbers could be too optimistic given the recent social upheavals and volatile economic growth that ensued.

Part of this pattern follows directly from the process of "bunching up above 1.25$ a day and just below 2$ a day" occurring in East Asia and, to a lesser extent, in South Asia over the last two decades (Chen and Ravallion, 2010). These two regions have a relatively large population near the poverty line and hence most of the advances are projected to occur there. Latin America and the Caribbean, as well as the Middle East and North Africa, are richer and require stronger income growth to continuously reduce poverty. Sub-Saharan Africa, on the other hand, has a considerable proportion of the population far below the 2$ a day line in 2010 (with 48.47% poor at 1.25$ a day). It is facing a lower income elasticity and thus *requires exceptionally strong income growth* to make significant strides against poverty. As highlighted in the previous section, this heterogeneity in the income elasticity is mainly due to income differences.

We repeat the same exercise for the 1.25$ a day poverty line. Table B-8 and Figure B-5 in Appendix B show the results based on the estimates presented in Table B-7. The performance of our method is similar and, when compared to the World Bank approach, is just slightly less accurate for the 2010 baseline. The broad patterns are also similar but start from much lower poverty levels. It is worth noting that the gap between Sub-Saharan Africa and South Asia is even wider for extreme poverty. All regions are predicted to have a poverty headcount below 7% in 2030, except Sub-Saharan Africa where we project poverty levels to remain at about 35%. In 2030, it is likely that the great majority of the world's extremely poor population will live on the Sub-Saharan subcontinent.

What do these results imply for the post-2015 development agenda? We suggest that a new goal to *at least* halve the 2$ a day poverty level within 20 years should be the bare minimum if we want to ensure steady progress. It could be combined with a more ambitious goal for extreme poverty (1.25$ a day) and significant resources targeted at Sub-Saharan Africa and South Asia. Shifting the policy focus to a higher poverty line makes a lot of sense. For most of the developing world the 1.25$ a day poverty line will become nearly irrelevant. In fact, as long as incomes continue to grow, any absolute poverty measure will become less relevant over time when it is set too low and 2$ a day can hardly be described as generous. China, for example, recently updated its national poverty line from less than 1$ a day to about 1.80$ a day. Raising the headline poverty threshold ensures the goal remains relevant as time passes.

Even the lower bound of this poverty reduction goal would not necessarily be self-fulling. A lasting slump in the developed world coupled with the possibility of China entering a middle-income trap could make it a challenge to preserve historical income growth rates throughout the medium-term future. In addition, income growth does not need to be distribution-neutral and anti-poverty policies will be more successful if they are accompanied by an improving income distribution. In any case, we now have a baseline to calculate goals against and to asses counterfactual assumptions. This can inform discussions on the post-2015 development agenda and help to identify reasonable benchmarks.

5 Conclusion

In this paper we derive a new approach to decomposing the poverty headcount and show that this fractional response approach outperforms earlier linear approximations. Our main point is that the well-established non-linearity of the income and inequality elasticities of poverty arises primarily from the bounded nature of the poverty headcount. Once this inherent non-linearity is taken into account, we can derive an empirical approximation of the poverty decomposition that implies income and inequality (semi-) elasticities with desirable properties.

We use this new approach to estimate income and inequality (semi-)elasticities of poverty based on a large new data set. Fractional response models fit the data extremely well. We provide evidence that the average income elasticity is around two and the average inequality elasticity is about one and a half. However, since these two averages are not very informative, we show that differences in income and inequality levels create strong regional heterogeneity in the estimated elasticities and semi-elasticities. Studies based on linear approximations do not fully capture this heterogeneity. Compared to earlier results, our approach provides estimates that are often substantially different, very stable and considerably more accurate. This holds for a wide range of different combinations of income and inequality. While our approach restricts the nature of the unobserved heterogeneity (measurement differences), it requires no distributional assumptions other than a correctly specified conditional mean. In addition, we show that classical measurement error in income attenuates the elasticity estimates and outweighs systematic survey bias pointing in the opposite direction.

Functional form matters a lot when estimating poverty decompositions. Elasticities and semi-elasticities of poverty estimated with fractional response models have properties closely resembling those of their theoretical counterparts derived under the assumption of log normality. Moreover, we emphasize that semi-elasticities rather than elasticities are the policy relevant metric. This non-linearity also has direct implications for a reduced-form literature interested in the poverty effects of more ultimate determinants or more policy-oriented variables. The focus should be first on how a particular variable of interest leads to changes in mean incomes and changes in distribution and, only in a second step, on how these changes bring about different poverty outcomes. Only in this manner, the non-linearity of the poverty-income-inequality relationship is properly taken into account.

To further illustrate the potential of the fractional response approach, we provide poverty projections from 2010 until 2030 based on the simple assumption that average historical consumption growth continues into the medium-term future. We show that the regional landscape of poverty is likely to change dramatically over the next two decades. Two findings stand out in particular. First, poverty in Sub-Saharan Africa and South Asia will remain the overarching challenge in the twenty-first century. Second, in all other regions poverty is projected to fall rapidly, so that there is a strong rationale for setting the post-2015 development goals on the basis of the 2$ a day poverty line.

It is tempting to interpret our findings as evidence of the primacy of growth. Yet, we are by no means arguing that income growth is all that matters for poverty reduction. It is important to emphasize that the causal effect of any particular policy on aggregate household income and the relative distribution of income cannot be discerned from a decomposition exercise such as this. What it does is help to identify how a *given* change in average income or in distribution translates into poverty outcomes, and not how that change is brought about. Hence, the importance of institutions, trade and a host of other factors for poverty alleviation remains undiminished. There is a potentially large 'double dividend' to be reaped if growth can be achieved in combination with simultaneous reductions in inequality.

Other important questions remain open. More work is needed on identifying viable paths of poverty alleviation that actually combine redistribution with growth. Such analyses require a more sophisticated political economy of redistribution and poverty than currently available. In addition, the issue of statistical discrepancies between expenditure surveys, national accounts and consumption proxies, and what these discrepancies imply for the confidence we place into poverty estimates, is just beginning to be explored. In our view, these are exciting areas for future research.

References

Acemoglu, D. and S. Johnson (2005). Unbundling institutions. *Journal of Political Economy 113*(5), 949–95.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91*(5), 1369–1401.

Adams, R. H. (2004). Economic growth, inequality and poverty: Estimating the growth elasticity of poverty. *World Development 32*(12), 1989–2014.

Aitchison, J. and J. A. C. Brown (1957). *The Lognormal Distribution: With Special References to Its Uses in Economics*. Cambridge University Press.

Altonji, J. G. and R. L. Matzkin (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica 73*(4), 1053–1102.

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics 104*, 184 – 198.

Battistin, E., R. Blundell, and A. Lewbel (2009). Why is consumption more log normal than income? Gibrat's law revisited. *Journal of Political Economy 117*(6), 1140–1154.

Beck, T., A. Demirgüç-Kunt, and R. Levine (2010). Financial institutions and markets across countries and over time: The updated financial development and structure database. *World Bank Economic Review 24*(1), 77–92.

Besley, T. and R. Burgess (2003). Halving global poverty. *Journal of Economic Perspectives 17*(3), 3–22.

Bluhm, R. (2015). fhetprob: A fast qmle stata routine for fractional probit models with multiplicative heteroskedasticity. Forthcoming as UNU-MERIT working paper.

Blundell, R. W. and J. L. Powell (2004). Endogeneity in semiparametric binary response models. *Review of Economic Studies 71*(3), 655–679.

Bourguignon, F. (2003). The growth elasticity of poverty reduction: Explaining heterogeneity across countries and time periods. In T. S. Eicher and S. J. Turnovsky (Eds.), *Inequality and Growth: Theory and Policy Implications*, pp. 3–26. Cambridge, MA: MIT Press.

Bresson, F. (2009). On the estimation of growth and inequality elasticities of poverty with grouped data. *Review of Income and Wealth 55*(2), 266–302.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2 of *Handbook of Econometrics*, pp. 1247–1318. Elsevier.

Chambers, D. and S. Dhongde (2011). A non-parametric measure of poverty elasticity. *Review of Income and Wealth 57*(4), 683–703.

Chen, S. and M. Ravallion (2004). How have the world's poorest fared since the early 1980s? *The World Bank Research Observer 19*(2), 141–169.

Chen, S. and M. Ravallion (2010). The developing world is poorer than we thought, but no less successful in the fight against poverty. *Quarterly Journal of Economics 125*(4), 1577–1625.

Cook, D. O., R. Kieschnick, and B. D. McCullough (2008). Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance 15*(5), 860–867.

Datt, G. and M. Ravallion (1992). Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s. *Journal of Development Economics 38*(2), 275–295.

Dollar, D. and A. Kraay (2002). Growth is good for the poor. *Journal of Economic Growth 7*(3), 195–225.

Dollar, D. and A. Kraay (2004). Trade, growth, and poverty. *Economic Journal 114*(493), F22–F49.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics 84*(2), 755–776.

Eichengreen, B., D. Park, and K. Shin (2013, January). Growth slowdowns redux: New evidence on the middle-income trap. Working Paper 18673, National Bureau of Economic Research.

Engerman, S. L. and K. L. Sokoloff (1997). Factor endowments, institutions, and differential growth paths among new world economies. In S. Haber (Ed.), *How Latin America Fell Behind*, pp. 260–296. Stanford, CA: Stanford University Press.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica 52*(3), 681–700.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*(6), 1251–1271.

Kakwani, N. (1993). Poverty and economic growth with application to Côte d'Ivoire. *Review of Income* *and Wealth 39*(2), 121–139.

Kalwij, A. and A. Verschoor (2007). Not by growth alone: The role of the distribution of income in regional diversity in poverty reduction. *European Economic Review 51*(4), 805–829.

Klasen, S. and M. Misselhorn (2008). Determinants of the growth semi-elasticity of poverty reduction. Working paper, Ibero America Institute for Economic Research.

Kraay, A. (2006). When is growth pro-poor? Evidence from a panel of countries. *Journal of Development Economics 80*(1), 198–227.

Loudermilk, M. S. (2007). Estimation of fractional dependent variables in dynamic panel data models with an application to firm dividend policy. *Journal of Business & Economic Statistics 25*(4), 462–472.

MacCullagh, P. and J. A. Nelder (1989). *Generalized linear models* (2nd ed.). Chapman & Hall/ CRC Press.

Milanovic, B. (2002). True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *Economic Journal 112*(476), 51–92.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*(1), 69–85.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*(1), 1–32.

Papke, L. E. and J. M. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics 11*(6), 619–632.

Papke, L. E. and J. M. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*(1–2), 121–133.

Ram, R. (2013). Income elasticity of poverty in developing countries: updated estimates from new data. *Applied Economics Letters 20*(6), 554–558.

Ramalho, E. A., J. J. Ramalho, and J. M. Murteira (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys 25*(1), 19–68.

Ravallion, M. (2001). Growth, inequality and poverty: Looking beyond averages. *World Development 29*(11), 1803–1815.

Ravallion, M. and S. Chen (1997). What can new survey data tell us about recent changes in distribution and poverty? *World Bank Economic Review 11*(2), 357–382.

Rivers, D. and Q. H. Vuong (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics 39*(3), 347–366.

Sachs, J. D. and A. Warner (1995). Economic reform and the process of global integration. *Brookings Papers on Economic Activity 1995*(1), 1–118.

Sala-i-Martin, X. (2006). The world distribution of income: Falling poverty and ... convergence, period. *Quarterly Journal of Economics 121*(2), 351–397.

Tiwari, A. and F. Palm (2011). Nonlinear panel data models with expected a posteriori values of correlated random effects. Working Paper 2011/13, CREPP HEC-Management School, University of Liège.

Wacziarg, R. and K. H. Welch (2008). Trade liberalization and growth: New evidence. *World Bank Economic Review 22*(2), 187–231.

Wooldridge, J. M. (2010a, May). Correlated random effects models with unbalanced panels. Manuscript.

Wooldridge, J. M. (2010b). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press.

Wooldridge, J. M. (2012, June). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. Manuscript.

Young, A. (2011, December). The gini coefficient for a mixture of ln-normal populations. Manuscript. The London School of Economics and Political Science, London, UK.

Appendix A

Let the poverty line (z) be fixed and assume the poverty headcount is described by a two-parameter distribution, so that $H(\bar{y}_t/z, \sigma_t) = H(\bar{y}_t, \sigma_t) = H_t$. A Taylor linearization of $H(\cdot)$ about $(\bar{y}_t, \sigma_t)$ gives

$$H(\bar{y}_t + \mathrm{d}\bar{y}_t, \sigma_t + \mathrm{d}\sigma_t) = H(\bar{y}_t, \sigma_t) + \frac{\partial H_t}{\partial \bar{y}_t}\mathrm{d}\bar{y}_t + \frac{\partial H_t}{\partial \sigma_t}\mathrm{d}\sigma_t + \xi_t \tag{A-1}$$

where $\mathrm{d}x$ denotes a differential change in x, and ξ_t is a second-order remainder. This is easily extended to allow for a vector of Lorenz curve parameters as in Kakwani (1993).

Subtracting $H(\bar{y}_t, \sigma_t)$ from both sides, dropping the remainder by approximation, dividing through by H_t (provided $H_t > 0$), and multiplying the first (second) term by $\bar{y}_t/\bar{y}_t$ (σ_t/σ_t), we arrive at eq. (4) from the main text:

$$\frac{\mathrm{d}H_t}{H_t} \approx \left(\frac{\partial H_t}{\partial \bar{y}_t}\frac{\bar{y}_t}{H_t}\right)\frac{\mathrm{d}\bar{y}_t}{\bar{y}_t} + \left(\frac{\partial H_t}{\partial \sigma_t}\frac{\sigma_t}{H_t}\right)\frac{\mathrm{d}\sigma_t}{\sigma_t} = \varepsilon_t^{H\bar{y}}\frac{\mathrm{d}\bar{y}_t}{\bar{y}_t} + \varepsilon_t^{H\sigma}\frac{\mathrm{d}\sigma_t}{\sigma_t}. \tag{A-2}$$

If we do not divide by H_t, we get a decomposition of the (non-relative) change of poverty in terms of income and inequality semi-elasticities (as in Klasen and Misselhorn, 2008).

Similar steps starting from $H(\bar{y}_t, G_t)$ lead to a decomposition in terms of mean income and Gini. Using the chain rule for elasticities, an expression for the Gini elasticity is

$$\varepsilon_t^{HG} = \varepsilon_t^{H\sigma}\left(\frac{\mathrm{d}G_t}{\mathrm{d}\sigma_t}\frac{\sigma_t}{G_t}\right)^{-1} \tag{A-3}$$

enabling us to write

$$\frac{\mathrm{d}H_t}{H_t} \approx \varepsilon_t^{H\bar{y}}\frac{\mathrm{d}\bar{y}_t}{\bar{y}_t} + \varepsilon_t^{HG}\frac{\mathrm{d}G_t}{G_t} = \varepsilon_t^{H\bar{y}}\frac{\mathrm{d}\bar{y}_t}{\bar{y}_t} + \varepsilon_t^{H\sigma}\left(\frac{\mathrm{d}G_t}{\mathrm{d}\sigma_t}\frac{\sigma_t}{G_t}\right)^{-1}\frac{\mathrm{d}G_t}{G_t} \tag{A-4}$$

where eqs. (2) and (3) give $\varepsilon_t^{H\bar{y}}$ and $\varepsilon_t^{H\sigma}$ under log normality, but we still need an expression for $\mathrm{d}G_t/\mathrm{d}\sigma_t$ to get an explicit formula for ε_t^{HG}.

Even though we restricted our attention to one inequality parameter, the results thus far are quite general. Now if we also assume log normality, we arrive at an explicit form for the Gini elasticity. Using $\sigma_t = \sqrt{2}\Phi^{-1}(G_t/2 + 1/2)$, we have

$$\frac{\mathrm{d}G_t}{\mathrm{d}\sigma_t} = \frac{\mathrm{d}[2\Phi\left(\sigma_t/\sqrt{2}\right) - 1]}{\mathrm{d}\sigma_t} = \sqrt{2}\phi\left(\frac{\sigma_t}{\sqrt{2}}\right). \tag{A-5}$$

Inverting and substituting eq. (A-5) together with eq. (3) from the main text into eq. (A-3) gives the Gini elasticity

$$\varepsilon_t^{HG} = \left(\frac{\ln(\bar{y}_t/z)}{\sigma_t} + \frac{1}{2}\sigma_t\right)\left(\frac{\sigma_t}{G_t}\sqrt{2}\phi\left(\frac{\sigma_t}{\sqrt{2}}\right)\right)^{-1}\lambda\left(\frac{-\ln(\bar{y}_t/z)}{\sigma_t} + \frac{1}{2}\sigma_t\right) \tag{A-6}$$

where $\sigma_t = \sqrt{2}\Phi^{-1}(G_t/2 + 1/2)$. This result corrects for the missing σ_t/G_t in Kalwij and Verschoor (2007, p. 824). The Gini semi-elasticity (η_t^{HG}) is just eq. (A-6) with $\phi(\cdot)$ replacing $\lambda(\cdot)$. Clearly, both the Gini elasticity and the Gini semi-elasticity are highly non-linear functions, as illustrated in Figure 2.

Appendix B

Figure B-1 – Transformed headcount (2$ a day) and log mean income, by region

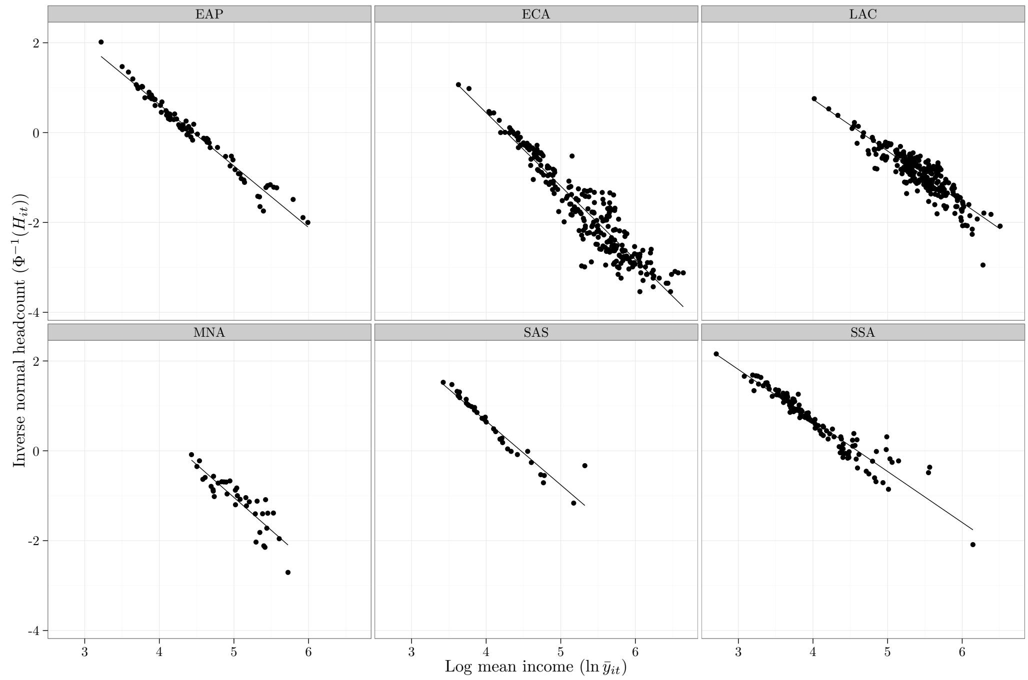


Figure B-2 – Transformed headcount (2$ a day) and log Gini, by region

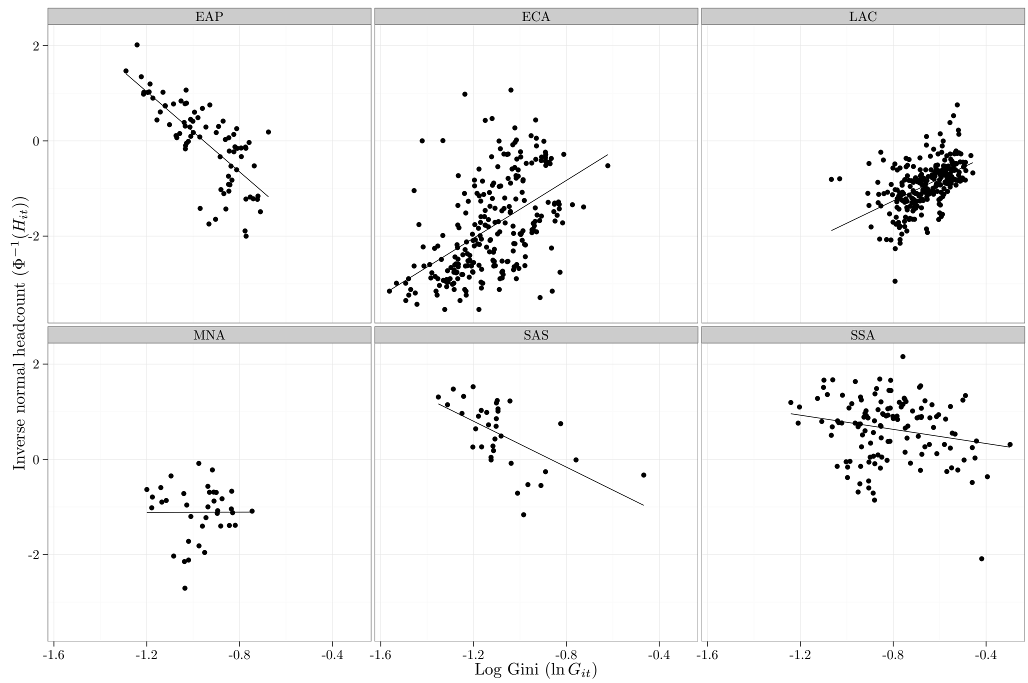


Table B-1 – Summary statistics by region (unweighted)

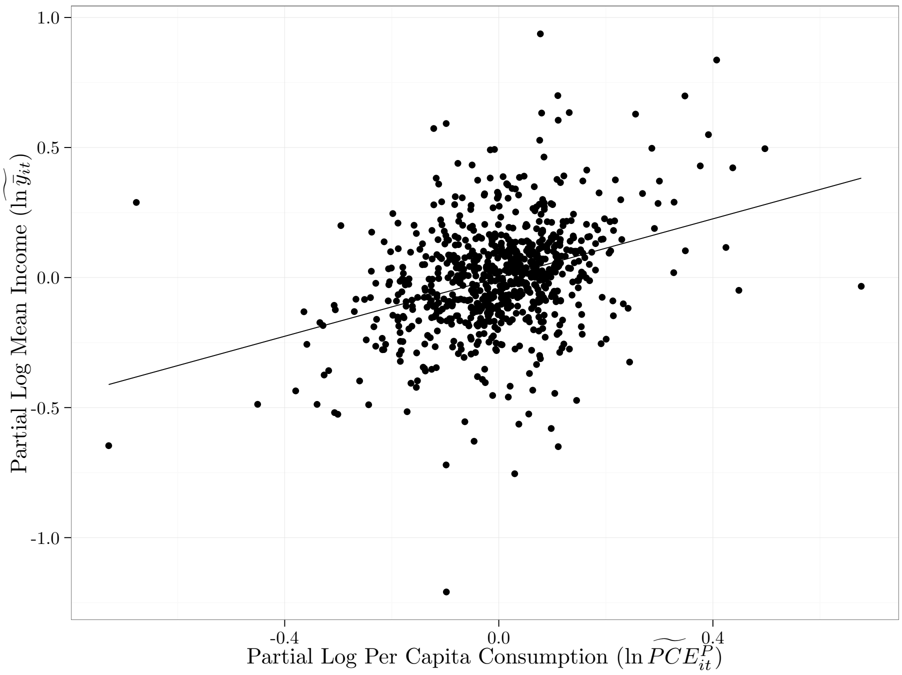
Variable	Mean	Standard Deviation	Min	Max
	East Asia and Pacific (N=80)			
H_{it} – Headcount (2$)	0.502	0.267	0.023	0.978
G_{it} – Gini coefficient	0.392	0.058	0.275	0.509
$\bar{y}_{it}$ – Mean income or expenditure	107.86	78.39	25.02	399.76
	Eastern Europe and Central Asia (N=254)			
H_{it} – Headcount (2$)	0.110	0.169	0.000	0.857
G_{it} – Gini coefficient	0.330	0.056	0.210	0.537
$\bar{y}_{it}$ – Mean income or expenditure	251.99	136.11	37.66	766.78
	Latin America and Caribbean (N=274)			
H_{it} – Headcount (2$)	0.204	0.122	0.002	0.775
G_{it} – Gini coefficient	0.523	0.054	0.344	0.633
$\bar{y}_{it}$ – Mean income or expenditure	246.63	90.55	55.53	671.04
	Middle East and North Africa (N=37)			
H_{it} – Headcount (2$)	0.166	0.111	0.003	0.466
G_{it} – Gini coefficient	0.380	0.042	0.301	0.474
$\bar{y}_{it}$ – Mean income or expenditure	165.26	56.59	84.02	306.33
	South Asia (N=35)			
H_{it} – Headcount (2$)	0.672	0.226	0.122	0.936
G_{it} – Gini coefficient	0.343	0.067	0.259	0.627
$\bar{y}_{it}$ – Mean income or expenditure	67.78	39.20	30.71	204.98
	Sub-Saharan Africa (N=129)			
H_{it} – Headcount (2$)	0.708	0.202	0.018	0.985
G_{it} – Gini coefficient	0.453	0.087	0.289	0.743
$\bar{y}_{it}$ – Mean income or expenditure	67.62	54.04	14.93	465.80

Notes: Mean income or expenditure in $ per month. 809 observations, 124 countries in total, unbalanced sample from 1981 to 2010.

List B-1 – Included countries (number of surveys)

Albania (5), Algeria (2), Angola (2), Argentina (21), Armenia (11), Azerbaijan (3), Bangladesh (8), Belarus (14), Belize (7), Benin (1), Bhutan (2), Bolivia, Plurinational State of (11), Bosnia and Herzegovina (3), Botswana (2), Brazil (26), Bulgaria (7), Burkina Faso (4), Burundi (3), Cambodia (5), Cameroon (3), Cape Verde (1), Central African Rep. (3), Chad (1), Chile (10), China (16), Colombia (14), Comoros (1), Congo, Dem. Rep. of (1), Congo, Rep. of (1), Costa Rica (23), Cote D'Ivoire (9), Croatia (7), Czech Rep. (2), Djibouti (1), Dominican Rep. (16), Ecuador (12), Egypt (5), El Salvador (14), Estonia (9), Ethiopia (4), Fiji (2), Gabon (1), Gambia (2), Georgia (14), Ghana (5), Guatemala (8), Guinea (4), Guinea-Bissau (2), Guyana (2), Haiti (1), Honduras (20), Hungary (10), India (5), Indonesia (13), Iran, Islamic Rep. of (5), Iraq (1), Jamaica (7), Jordan (7), Kazakhstan (11), Kenya (4), Kyrgyzstan (10), Lao People's Dem. Rep. (4), Latvia (11), Lesotho (4), Liberia (1), Lithuania (9), Macedonia, Rep. of (10), Madagascar (6), Malawi (3), Malaysia (9), Maldives (2), Mali (4), Mauritania (6), Mexico (11), Micronesia, Federated States of (1), Moldova, Rep. of (15), Montenegro (4), Morocco (5), Mozambique (3), Namibia (2), Nepal (4), Nicaragua (4), Niger (4), Nigeria (5), Pakistan (8), Palestinian Territory, Occupied (2), Panama (13), Papua New Guinea (1), Paraguay (14), Peru (16), Philippines (9), Poland (18), Romania (15), Russian Federation (13), Rwanda (3), Saint Lucia (1), Sao Tome and Principe (1), Senegal (4), Serbia (9), Seychelles (1), Sierra Leone (1), Slovakia (7), Slovenia (6), South Africa (5), Sri Lanka (6), Sudan (1), Suriname (1), Swaziland (3), Syrian Arab Rep. (1), Tajikistan (5), Tanzania, United Rep. of (3), Thailand (14), Timor-Leste (2), Togo (1), Trinidad and Tobago (2), Tunisia (6), Turkey (11), Turkmenistan (3), Uganda (7), Ukraine (13), Uruguay (7), Venezuela, Bolivarian Rep. of (13), Vietnam (6), Yemen (2), Zambia (7).

Figure B-3 – Partial regression plot – first stage



Notes: The figure plots two residual series, so that the plotted slope is identical to the slope of $\ln PCE^P_{it}$ in the first stage. On the x-axis: $\widetilde{\ln PCE^P_{it}} = \ln PCE^P_{it} - \mathbf{x}'_{1it}\hat{\boldsymbol{\beta}}_1 - \sum_{r=1}^{T} \delta_{T_i,r}\hat{\varphi}_{1r} - \sum_{r=1}^{T} \delta_{T_i,r}\bar{\mathbf{x}}'_i\hat{\boldsymbol{\theta}}_{1r}$. On the y-axis: $\widetilde{\ln \bar{y}_{it}} = \ln \bar{y}_{it} - \mathbf{x}'_{1it}\hat{\boldsymbol{\beta}}_1 - \sum_{r=1}^{T} \delta_{T_i,r}\hat{\varphi}_{1r} - \sum_{r=1}^{T} \delta_{T_i,r}\bar{\mathbf{x}}'_i\hat{\boldsymbol{\theta}}_{1r}$. In both cases, $\mathbf{x}'_{1it}$ includes only the log of Gini but $\bar{\mathbf{x}}'_i$ contains the time averages of $\ln G_{it}$ and $\ln PCE^P_{it}$. Both regressions also contain survey type and time dummies, as well as their time averages.

Figure B-4 – Inequality changes and income growth, 1981–2010

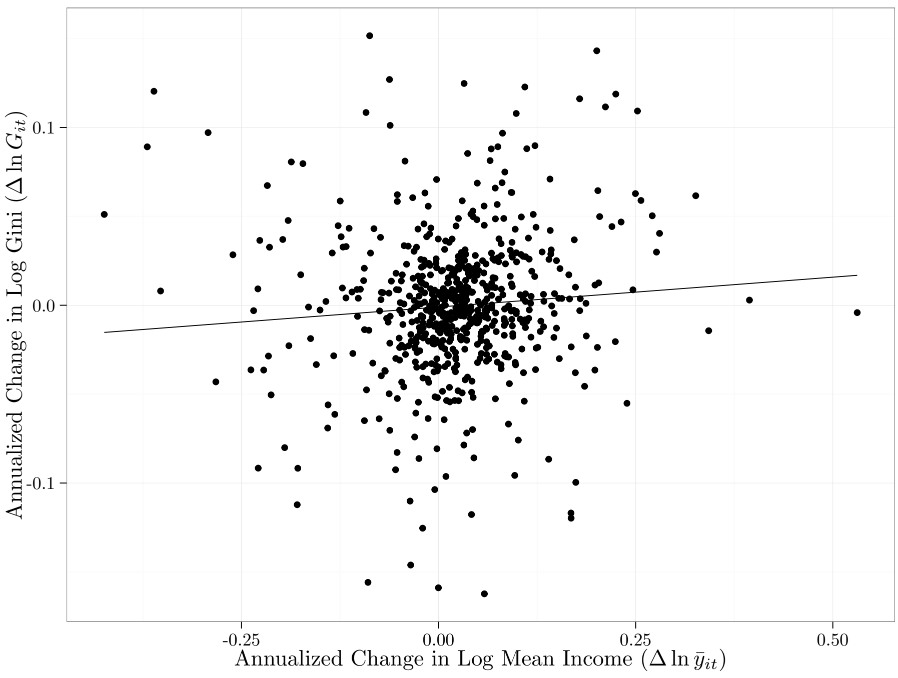


Table B-2 – Linear models – Dependent variable: $\ln H_{it}$, 2$ a day

	OLS			Two-Step GMM		
	(1)	(2)	(3)	(4)	(5)	(6)
	Within	Differences	Differences	Differences	Differences	Differences
	R+C '97	R+C '97	Bourg. '03	R+C '97	Bourg. '03	K+V '07
$\Delta \ln \bar{y}_{it}$		-1.895	-0.268	-2.028	2.046	-0.362
		(0.170)	(0.617)	(0.271)	(1.043)	(3.216)
$\Delta \ln \bar{y}_{it} \times \ln G_{i,t-1}$			1.108		3.445	2.097
			(0.671)		(1.192)	(2.315)
$\Delta \ln \bar{y}_{it} \times \ln(\bar{y}_{i,t-1}/z)$			-0.552		-0.995	-0.517
			(0.179)		(0.258)	(0.785)
$\Delta \ln G_{it}$		2.336	-0.527	1.664	1.257	-8.222
		(0.311)	(1.449)	(1.008)	(4.127)	(11.185)
$\Delta \ln G_{it} \times \ln G_{i,t-1}$			-1.769		-1.416	-8.164
			(1.586)		(3.929)	(8.296)
$\Delta \ln G_{it} \times \ln(\bar{y}_{i,t-1}/z)$			1.261		-0.315	-1.382
			(0.427)		(1.172)	(1.996)
$\ln \bar{y}_{it}$	-2.114					
	(0.204)					
$\ln G_{it}$	3.024					
	(0.409)					
$\ln G_{i,t-1}$						-0.129
						(0.134)
$\ln(\bar{y}_{i,t-1}/z)$						-0.023
$\bar{\varepsilon}^{H\bar{y}}$	-2.114	-1.895	-1.755	-2.028	-1.905	-2.684
$\bar{\varepsilon}^{HG}$	3.024	2.336	2.201	1.664	2.206	-2.345
$N \times \bar{T}$	648	648	648	641	641	641
N	104	104	104	102	102	102
Hansen's J (p-val.)	–	–	–	0.0418	0.579	0.639

Notes: All standard errors are robust to clustering at the country-level. The GMM results are estimated using two-step efficient GMM. Model (4) uses as instruments ΔPCE_{it}, $PCE_{i,t-1}$, $\ln \bar{y}_{i,t-1}$ and $\ln G_{i,t-1}$. Model (5) uses as instruments ΔPCE_{it}, $PCE_{i,t-1}$, $\Delta PCE_{it} \times \ln G_{i,t-1}$, $\Delta PCE_{it} \times \ln(\bar{y}_{i,t-1}/z)$, $\ln \bar{y}_{i,t-1}$, $\ln \bar{y}_{i,t-1} \times \ln G_{i,t-1}$, $\ln \bar{y}_{i,t-1} \times \ln(\bar{y}_{i,t-1}/z)$, $\ln G_{i,t-1}$ and $\ln G_{i,t-1} \times \ln G_{i,t-1}$. Model (6) uses the same instruments as model (5) but $\ln \bar{y}_{i,t-1}$ and $\ln G_{i,t-1}$ instrument for themselves. All models include a constant (not shown) and model (1) includes a time trend (not shown). Models (2) and (4) are similar to Ravallion and Chen (1997) but we update their approach by also including the Gini as in Adams (2004), models (3) and (5) are similar to the "improved standard model 2" in Bourguignon (2003), and model (6) is in the spirit of the preferred specification in Kalwij and Verschoor (2007). The latter also use the annualized log difference of the population ($\Delta \ln pop_{it}$) as an instrument and rely on real GDP per capita instead of real per capita consumption. A first-stage F-test shows that $\Delta \ln pop_{it}$ is an extremely weak instrument. Kalwij and Verschoor (2007) also use interactions of lagged inequality and lagged income with regional dummies as instruments. However, first stage diagnostics suggest a weak IV problem (the F-stat with regional dummy interactions is always lower than without) and thus we opt for a simpler instrument set. Further, in model (5) and equation (5) we do not include the lagged levels of income and inequality. Model (6) includes them for comparison with Kalwij and Verschoor (2007).

Table B-3 – Predicted regional income semi-elasticities, preferred specification

	Time period				
	1981–1989	1990–1994	1995–1999	2000–2004	2005–2010
East Asia and Pacific	-0.568	-0.573	-0.585	-0.583	-0.552
	(0.034)	(0.036)	(0.046)	(0.042)	(0.051)
Eastern Europe and Central Asia	-0.031	-0.214	-0.260	-0.225	-0.134
	(0.008)	(0.015)	(0.020)	(0.015)	(0.010)
Latin America and Caribbean	-0.374	-0.348	-0.334	-0.355	-0.194
	(0.028)	(0.025)	(0.024)	(0.026)	(0.013)
Middle East and North Africa	-0.405	-0.422	-0.447	-0.463	-0.313
	(0.034)	(0.037)	(0.042)	(0.043)	(0.024)
South Asia	-0.418	-0.458	-0.526	-0.572	-0.585
	(0.023)	(0.019)	(0.022)	(0.036)	(0.044)
Sub-Saharan Africa	-0.532	-0.354	-0.353	-0.440	-0.459
	(0.024)	(0.020)	(0.020)	(0.015)	(0.015)

Notes: Standard errors obtained via a panel bootstrap using 999 replications. The predicted semi-elasticities are based on estimated APEs at each region/time-period mean of $\ln \bar{y}_{it}$ and $\ln G_{it}$.

Table B-4 – Predicted regional Gini semi-elasticities, preferred specification

	Time period				
	1981–1989	1990–1994	1995–1999	2000–2004	2005–2010
East Asia and Pacific	0.419	0.423	0.432	0.431	0.408
	(0.053)	(0.053)	(0.055)	(0.054)	(0.053)
Eastern Europe and Central Asia	0.023	0.158	0.192	0.166	0.099
	(0.007)	(0.015)	(0.019)	(0.017)	(0.012)
Latin America and Caribbean	0.276	0.257	0.247	0.262	0.143
	(0.046)	(0.043)	(0.043)	(0.045)	(0.029)
Middle East and North Africa	0.299	0.311	0.330	0.342	0.231
	(0.041)	(0.040)	(0.041)	(0.044)	(0.025)
South Asia	0.309	0.338	0.389	0.423	0.432
	(0.056)	(0.055)	(0.052)	(0.054)	(0.055)
Sub-Saharan Africa	0.393	0.261	0.261	0.325	0.339
	(0.050)	(0.037)	(0.040)	(0.042)	(0.043)

Notes: Standard errors obtained via a panel bootstrap using 999 replications. The predicted semi-elasticities are based on estimated APEs at each region/time-period mean of $\ln \bar{y}_{it}$ and $\ln G_{it}$.

Table B-5 – Growth in personal consumption expenditures per capita (in %)

Region	Last 5 years	Last 10 years	Last 15 years	Last 20 years
East Asia and Pacific	6.843	5.962	5.647	5.967
	(1.041)	(0.782)	(0.837)	(0.757)
Europe and Central Asia	4.532	6.033	4.793	2.856
	(1.039)	(1.063)	(0.496)	(0.423)
Latin America and the Caribbean	3.364	2.399	2.222	2.267
	(0.746)	(0.303)	(0.172)	(0.151)
Middle East and North Africa	2.705	3.778	2.911	2.499
	(0.634)	(0.560)	(0.370)	(0.349)
South Asia	5.563	4.684	4.123	3.636
	(0.865)	(0.556)	(0.537)	(0.438)
Sub-Saharan Africa	1.710	2.765	2.338	1.742
	(1.599)	(0.682)	(0.577)	(0.414)
$N \times \bar{T}$	615	1222	1795	2332
N	123	123	123	123

Notes: Cluster robust standard errors in parentheses. Regional means are population weighted. Only the third column is relevant for the projections.

Table B-6 – Fractional probit models (QMLE) – Dependent variable: H_{it}, 2$ a day

	(1) Institutions		(2) Human Capital		(3) Credit		(4) Trade	
	H_{it}	APEs	H_{it}	APEs	H_{it}	APEs	H_{it}	APEs
$\ln \bar{y}_{it}$	-0.888	-0.285	-0.878	-0.284	-0.950	-0.289	-0.708	-0.302
	(0.050)	(0.012)	(0.060)	(0.011)	(0.036)	(0.009)	(0.032)	(0.012)
$\ln G_{it}$	0.779	0.250	0.805	0.261	0.765	0.233	0.581	0.248
	(0.107)	(0.028)	(0.104)	(0.027)	(0.102)	(0.027)	(0.097)	(0.033)
Executive Constraints	0.005	0.001						
	(0.005)	(0.001)						
Year of Schooling			-0.002	-0.001				
			(0.017)	(0.006)				
Private Credit / GDP					-0.007	-0.002		
					(0.040)	(0.012)		
Trade Openness							0.005	0.002
							(0.017)	(0.007)
Scale factor	0.321		0.324		0.304		0.426	
$N \times \bar{T}$	678		705		697		385	
N	85		87		93		81	
AIC	894.8		914.1		887.6		552.5	
$\ln \mathcal{L}$	-276.4		-286.1		-282.8		-163.2	
$\sqrt{MSE}$	0.0203		0.0211		0.0201		0.0233	

Notes: The estimation samples are smaller due to less data coverage and all observations with $T_i = 1$ are not used in estimation. All models include time averages (CRE), time dummies, survey dummies, panel size dummies and interactions between the panel size dummies and the time averages (CRE). The time averages are recomputed for each sample size. The coefficients of the time average of the survey dummies and time effects are constrained to be equal across the panel sizes. The variance equation depends on the panel size. The standard errors of the coefficients are robust to clustering at the country level and the standard errors of the APEs are computed via the delta method. Data on *Executive Constraints* is from the Polity IV database. Human capital is measured as *Total Years of Schooling* from Barro and Lee (2013). We linearly interpolate the quinquennial data to an annual series. Financial development measured as *Private Credit / GDP* is from Beck, Demirgüç-Kunt, and Levine (2010). De jure *Trade Openness* is the binary measure developed by Sachs and Warner (1995) and extended by Wacziarg and Welch (2008).

Table B-7 – Fractional probit models (QMLE) – Dependent variable: H_{it}, 1.25$ a day

	(1) Regular		(2) Unbalanced		(3) Unbalanced + Two-Step	
	H_{it}	APEs	H_{it}	APEs	H_{it}	APEs
$\ln \bar{y}_{it}$	-1.212	-0.216	-0.668	-0.218	-0.800	-0.263
	(0.056)	(0.010)	(0.038)	(0.008)	(0.180)	(0.034)
$\ln G_{it}$	1.238	0.221	0.726	0.237	0.714	0.235
	(0.121)	(0.022)	(0.074)	(0.020)	(0.180)	(0.032)
$\hat{\nu}_{it}$					0.104	
					(0.104)	
CRE (Corr. Rand. Effects)	Yes		Yes		Yes	
Survey type dummies	Yes		Yes		Yes	
Time dummies	Yes		Yes		Yes	
Panel size dummies	No		Yes		Yes	
Panel size dummies × CRE	No		Yes		Yes	
Variance equation	No		Yes		Yes	
Scale factor	0.179		0.326		0.329	
$N \times \bar{T}$	768		768		754	
N	103		103		102	
pseudo R^2	0.988		0.995		0.995	
$\ln \mathcal{L}$	-172.4		-244.7		-243.7	
$\sqrt{MSE}$	0.0339		0.0214		0.0220	

Notes: The 1.25$ a day sample is smaller as for 20 observation we only have data at the 2$ a day line. 21 observations with $T_i = 1$ are not used in estimation. In models (1) and (2), the standard errors of the coefficients are robust to clustering at the country level and the standard errors of the APEs are computed via the delta method. We include the time averages of the survey type and time dummies in (2) and (3), but constrain their coefficients to be equal across the panel sizes. The standard errors of the coefficients and the APEs in model (3) account for the first stage estimation step with a panel bootstrap using 999 bootstrap replications. The linear projection in the first stage uses $\ln PCE_{it}^{P}$ as an instrument for $\ln \bar{y}_{it}$. The first-stage cluster-robust F-statistic in (3) is 24.40. Model (3) also excludes West Bank and Gaza entirely (2 observations) and 12 observations from ECA countries pre-1990 for lack of PCE data.

Table B-8 – Poverty projections ($\hat{H}_{it} \times 100$), preferred specification, 1.25$ a day

	2010 Official	2010 Estimate	2015 Estimate	2020 Estimate	2025 Estimate	2030 Estimate
East Asia & Pacific	12.48	9.63	5.09	2.76	1.60	1.00
Europe & Central Asia	0.66	0.74	0.44	0.27	0.16	0.10
Latin America & the Caribbean	5.53	5.59	4.84	4.22	3.70	3.26
Middle East & North Africa	2.41	3.43	2.48	1.82	1.37	1.06
South Asia	31.03	33.81	23.89	16.09	10.53	6.89
Sub-Saharan Africa	48.47	46.87	42.95	39.84	37.20	34.86
World Total	20.63	20.44	15.82	12.66	10.58	9.27

Notes: Regional aggregates are based on the World Bank classification of low and middle income countries in 1990. High income countries in 1990 are assumed to have no poor. The projections are for 123 countries. West Bank and Gaza is excluded as we lack both PCE and population data.

Figure B-5 – Predicted regional poverty, 1.25$ a day, 2010 to 2030