Income Distribution and Economic Development: Insights from Machine Learning

Pushan Dutt*[a] Ilia Tsetlin[b]

INSEAD & ABFER INSEAD

March 2020

Abstract

We draw upon recent advances that combine causal inferences with machine learning, to show that poverty is the key income distribution measure that matters for development outcomes. In a predictive framework, we first show that LASSO chooses only the headcount measure of poverty from 37 income distribution measures in predicting schooling, institutional quality, and per capita income. Next, causal inferences with post-LASSO models indicate that poverty matters more strongly for development outcomes than does the Gini coefficient. Finally, instrumental variable estimates in conjunction with post-LASSO models show that compared to Gini, poverty is more strongly causally associated with schooling and per capita income, but not institutional quality. Our results question the literature's overwhelming focus on the Gini coefficient. At the least, our results imply that the causal link from inequality (as measured by Gini) to development outcomes is tenuous.

JEL Classification: D31; I32; O10

Keywords: poverty; inequality; income distribution; economic development

*Corresponding author: INSEAD, 1 Ayer Rajah Avenue, Singapore 138676; Tel: +65 6799 5498
a: INSEAD, 1 Ayer Rajah Avenue, Singapore 138676
b: INSEAD, 1 Ayer Rajah Avenue, Singapore 138676
E-mail addresses: Pushan.Dutt@insead.edu ; Ilia.Tsetlin@insead.edu

1. Introduction

Recent years have seen a renewed focus on inequality going by the extraordinary response to Thomas Piketty's "Capital in the Twenty-First Century." Piketty (2014) highlights that rising inequality in many advanced economies since 1980 is predominantly driven by the gains in income shares at the very top – the top 1%, the top 0.1%. This focus on inequality at the top stands in stark contrast to the rich literature relating inequality to developmental outcomes such as economic growth (Alesina and Rodrik, 1994; Persson and Tabellini, 1994), schooling (Galor, 2011; Galor, Moav and Vollrath, 2009), institutional quality (Glaeser Scheinkman, and Shleifer, 2003; Perotti, 1996) and democratization (Boix, 2003). Most work measures inequality using the Gini coefficient or, in some cases, the income share of the median quintile. Despite the availability of better quality datasets (Deininger and Squire, 1998) the literature has failed to reach a consensus on whether and how inequality matters for development outcomes. In contrast to previous findings that demonstrate a negative relationship between inequality and development, others find either a positive relationship (Forbes, 2000)[1] or a zero relationship between the two (Barro, 2000). Banerjee and Duflo (2003) highlight the non-linear relationship between inequality and growth to reconcile these divergent findings. At the same time, they are careful to acknowledge that these are correlations and that causality is hard to sort out.

Easterly (2007) takes causality seriously. Building on Engerman and Sokoloff (1997), Easterly (2007) uses agricultural endowments as an instrument for inequality (specifically, the abundance of land suitable for growing wheat relative to that suitable for growing sugarcane) to show that inequality is indeed causally related to lower per capita incomes. Easterly (2007) also identifies two channels via which inequality reduces per capita income. He demonstrates that

[1] Lee and Son (2016) find that with more updated data on Gini and a more comprehensive set of controls, inequality does reduce economic growth.

countries with higher inequality exhibit lower levels of human capital and poor institutional quality. What unifies all this work is the near-universal focus on the Gini coefficient as the summary statistic for inequality.[2] Banerjee and Duflo (2003), for instance, question the assumption that the Gini coefficient is the appropriate measure of inequality, suggesting that measures of poverty or interquartile range are equally valid candidates. Nevertheless, they proceed to present all results with the Gini coefficient.[3] Gimpelson and Triesman (2017) show that there is widespread ignorance and misperceptions of the Gini coefficient itself.

In this paper we extend the focus from the Gini coefficient to an array of measures of the overall income distribution. Given that there are multiple ways to measure income distribution, we start by adopting a prediction approach from machine learning. In particular, we use linear, high dimensional sparse (HDS) regression models in econometrics (see Belloni, Chernozhukov and Hansen, 2013, 2014a, for comprehensive overviews and Vapnik, 2013, for theoretical foundations) which allows for a large number of regressors, possibly much larger than the sample size, but imposes a sparsity restriction on the model. That is, these models assume only a subset of these regressors, are important for capturing the main features of the regression function. We first set aside inference, take a purely predictive approach and use elastic net (Zou and Hastie, 2005), LASSO (least absolute shrinkage and selection operator) and its variant, a post-LASSO method from Belloni and Chernozhukov (2013) to select from multiple income distribution measures. We find that from a pure prediction perspective, it is poverty that matters rather than any other distributional statistic in predicting development outcomes – per capita GDP, schooling, and institutions, used in Easterly (2007).

[2] Voitchovsky (2005) highlights that even when the mechanisms by which inequality affects economic growth differ, the empirical work almost exclusively relies on the Gini coefficient. Some work uses the share of the median quintile (e.g., Persson and Tabellini, 1994; Easterly, 2001) in addition to the Gini coefficient. Beck, Demirgüç-Kun and Ross (2007) is one of the few that examines Gini and income share of the poorest quintile (measures of relative inequality) and percentage of poor living on less than $1 a day (measure of absolute poverty). Lupu and Peterson (2011) is one of the few papers that go beyond the Gini coefficient emphasizing skewness and the detailed structure of the income distribution.

[3] An exception is Ravillion (2012) who highlights that initial poverty matters for growth and convergence.

Inequality, as measured by the Gini coefficient, is a measure of the relative disparities in levels of living standard while poverty encapsulates absolute levels of living – how many people fail to attain a certain predetermined consumption need (Ravallion, 2003). There are plausible reasons why poverty emerges as a more important factor for economic development. Poverty hurts human capital especially in the presence of credit constraints (Moav, 2005), poverty traps consign economies to low levels of underdevelopment (Mookherjee and Ray, 2003; Ghatak and Jiang, 2002), poverty allows the wealthy to subvert institutions (Glaeser et al, 2003) and by reducing productivity hurts incomes (Banerjee and Mullainathan, 2008). Not having enough money for satisfying basic needs (which defines poverty) is very different from having an income, which is somewhat lower than the income of other people, which is related to the Gini coefficient. Ravillion (2012) shows that poverty acts directly as a constraint on growth *and* makes it harder to achieve any given proportionate impact on poverty through growth.

Next, we shift our focus to causal inference and examine the relative importance of inequality vs. poverty for development outcomes. This is challenging since it can easily be argued that both poverty and the Gini coefficient are also development outcomes and affected by schooling, institutions, and per capita income. While our income distribution measures are calculated for the year 1988 and the development outcomes measured at least a decade later, we may still have an omitted variable problem. Therefore, we start by recognizing that a multiplicity of variables can potentially impact both development outcomes and income distribution, which makes it challenging to answer causal questions in a cross-country context confidently. As Sala-i-Martin, Doppelhofer, and Miller (2004) argue, even theory is not very helpful in making sense of the empirical evidence. Multiple models exist that “predict” that a particular variable (e.g., distortions, disease burden, property rights, degree of monopoly power, demographics, etc.) matters for economic development. However, these theories need not be mutually exclusive, and with a small number of observations (number of countries for which data are available), we do not have a large enough sample size to evaluate the relative importance of the set of potential regressors. Essentially, the number of parameters that can be considered (*p)* is large relative to the sample size (*n*), i.e., $p >> n$. We again draw on recent extensions of machine learning to causal inference that allow for dimensional reduction and inference when the number of parameters is large. We apply the post-Lasso double selection method of Belloni, Chernozhukov, and Hansen (2014b) that systematically selects from a large set of 67 potential confounders. In a first-stage, we use a Lasso-type procedure for variable selection to predict both the dependent variable (development outcome in our case) and the main independent variable (s) (poverty or Gini or both in our case).[4] We apply standard LASSO to choose a subset of variables used in Sala-i-Martin (1997) and Sala-i-Martin et al. (2004). In the second stage, we estimate the effect of interest by the linear regression of the outcome variable(s) on the main independent variable(s) and the union of the set of variables selected in the variable selection steps. This post-Lasso double selection procedure also demonstrates that poverty matters more than Gini for schooling, institutional quality, and per capita income. The procedure also suggests that dropping poverty results in a serious omitted variable bias while the Gini coefficient is not an important predictor for either development outcomes or for poverty. Overall, our results show that for the most part, poverty has a stronger influence on schooling, institutions, and income, and that including poverty makes the Gini coefficient results weaker and/or insignificant.

Next, since the set of omitted variables is potentially infinite, we switch to an instrumental variable strategy in conjunction with machine-learning techniques. We employ the same land-endowment based instruments as in Easterly (2007) and show that these instruments strongly impact poverty and that the land-endowment instrument affects development outcomes

[4] To facilitate comparison, we use Easterly (2007) as our baseline, and use poverty in lieu of and in conjunction with Gini as the key income distribution measure.

through poverty rather than inequality. We take as given that these are valid instruments and we first show that simply adding the (uninstrumented) poverty measure to the regressions where inequality is instrumented is sufficient to make inequality insignificant. Combining the double-selection procedure of Belloni, Chernozhukov and Hansen (2014b) and the instruments from Easterly (2007) we subsequently show that poverty matters for schooling and per capita GDP, and not the Gini coefficient. At the same time, neither measure of income distribution matters for institutional quality.

Overall, our instrumental variable results suggest that even in a cross-country setting, the causal link from inequality to development outcomes is less robust than widely accepted in the literature (see Benabou, 2000).[5] At best, our results show that poverty rather than inequality matters more – for many countries, the focus, perhaps, should be on inequality at the bottom rather than inequality at the top. If it is poverty that matters, this has very different implications for redistributive measures adopted by policymakers. In essence, it calls for a focus on poverty alleviation rather than distributing income from the top 1% towards the middle class. Despite the use of instruments and reliance on machine learning estimators designed for causal inferences, we are aware that sorting out causality in cross-sectional regressions is a hard task. Therefore, a conservative interpretation of our findings is that the exclusion restrictions that commonly used instruments affect development outcomes only through the Gini coefficient are questionable.

The rest of the paper is organized as follows. In Section 2, we discuss various measures that summarize the income distribution and development outcomes used in the prior literature. Section 3 uses a predictive framework to assess which measures best predict development outcomes. In Section 4, we provide a short example to highlight that an absolute measure of

[5] A recent paper by Sarsons (2015) makes a similar point on the use of rainfall as an instrument for income shocks. She shows that while rainfall is plausibly exogenous, it affects civil conflict through a variety of channels and not just via income.

poverty and the Gini coefficient can diverge in various ways, and a priori it is not clear which income distribution is preferable. In Section 5, we move to causal inference and use a LASSO based double-selection methodology to infer the role of poverty for economic development. We also combine the double-selection methodology with the instrumenting strategy in Easterly (2007) to again show that poverty matters more than the Gini coefficient. Section 6 concludes.

2. Income Distribution and Development Outcomes

There are multiple ways to summarize income distribution. We can think in terms of multiple inequality measures, shares of different deciles or quintiles, multiple poverty measures, absolute vs. relative poverty lines, etc. For instance, there are at least two close weighted variants of the Gini coefficient – the Mehran index which is more sensitive to changes in the lower end of the distribution, when compared to the Gini index, and the Piesch index which is more sensitive to changes in the upper end of the distribution (see Yitzhaki, 1983). Alternate inequality measures exist as well, such as the family of Generalized Entropy measures, and the Atkinson measure, which allows for varying sensitivity to inequalities in different parts of the income distribution. Similarly, if we focus exclusively on the bottom of the income distribution and the headcount measure of poverty, we still have a choice in terms of poverty lines. The two commonly used poverty measures are headcounts based on the World Bank poverty lines of $1.25 a day and $2 a day. The ease of interpretation of the headcount measures accounts for much of their popularity.

Given the multiplicity of such measures, we use the Milanovic (2002) database on world income distribution to construct a comprehensive set, all of which capture varying aspects of the income distribution. There are multiple advantages to using this database. First, it is based on household surveys that permit richer and more accurate measures of income distribution *within* countries, by deciles in this case. Second, the surveys also provide information on mean incomes within deciles, which is a far more accurate measure of household incomes and expenditures as compared to a crude measure such as per capita GDP. GDP, for instance, includes undistributed profits or increase in stocks, which may be orthogonal to the welfare of the population. The data on mean incomes are adjusted for differences in purchasing power to facilitate comparability across countries. Finally, this database combines the internationally comparable poverty monitoring database (PovcalNet) compiled by the World Bank (see Chen and Ravallion, 2010, for more details) and the Luxembourg Income Study (LIS) which allows for the inclusion of advanced economies. For all income distribution measures, we use data from the year 1988, the earliest year for which the data are available.[6]

2.1 Measures of Income Distribution

We use this data to construct the Gini coefficient, the Mehran index, and the Piesch index for the year 1988. The standard Gini index measures twice the surface between the Lorenz curve, which maps the cumulative income share on the vertical axis against the distribution of the population on the horizontal axis, and the line of equal distribution. The Mehran and the Piesch indices are similar to the Gini index, except that they employ weights. For the Mehran index, the difference between the ordinate of the line of perfect equality and the ordinate of the Lorenz curve is weighted by $1 - p_i$, where p_i is the horizontal coordinate of the Lorenz curve. This makes the Mehran index relatively more sensitive to changes in the lower end of the distribution when compared to the Gini. In the Piesch index the weighting factor is p_i, making it relatively more sensitive to changes in the upper end of the distribution as compared to the Gini.

Next, we construct three generalized entropy measures. The generalized entropy class of measures are given by $GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{\alpha} - 1\right]$ where N is the number of

[6] The data for this study are available from the corresponding author upon request.

individuals in the sample, y_i is the income of individual i, $i \in (1, 2,..., N)$, $\bar{y}$ is mean income, and $\alpha \neq 0,1$. Measures from the generalized entropy class are sensitive to changes at the lower end of the distribution for values of α close to zero, equally sensitive to changes across the distribution for α equal to one, and sensitive to changes at the higher end of the distribution for higher values.We use the three most commonly used values of α, namely α = 0, 1, 2 (Haughton and Khandker, 2009). $GE(0) = \left[\frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{\bar{y}}{y_i}\right)\right]$, is known as Theil's L, and sometimes referred to as the mean log deviation. $GE(1) = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)\log\left(\frac{y_i}{\bar{y}}\right)\right]$ is the Theil's T index, and GE(2) is half of the coefficient of variation.

The next class of inequality measures we construct are those by Atkinson (1987). For a weighting parameter ε, which captures aversion to inequality, the Atkinson class is defined as $A(\varepsilon) = 1 - \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{1-\varepsilon}\right]^{1/(1-\varepsilon)}$. We set the weighting parameter ε at three most commonly used values 0.5, 1, and 2 (Haughton and Khandker, 2009). Additionally, we included the coefficient of variation, the relative mean deviation and the standard deviation of log income. Finally, we use each of 10 decile shares in income as measures of income distribution.

Next, we construct multiple poverty measures again for the year 1988. We base our first poverty measures on the widely used World Bank benchmarks of $1.25 and $2.0 a day. For each decile, we define two dummy variables at the decile-county level, which take the value 1 if the mean annual income of the decile in a particular country is less than $456.25 and $730 respectively (1.25 a day* 365 days and 2.0 a day*365 days). The $456.25 ($730) cut-off corresponds to the poverty measure of $1.25 ($2.0) a day. Summing up these deciles by country, gives us our headcount poverty measures, $Pov^{1.25}$ and $Pov^{2.0}$ as the percentage of population with incomes below $1.25 a day and $2.0 dollars a day.[7] The correlation between

[7] Lacking more detailed data on income distributions, our poverty measures assume that all people within each decile (data point) have the same income. While this may bias the overall poverty measure, the direction of bias is not obvious a priori.

our measure and the widely reported headcount measures from the World Bank and available from the World Development Indicators is 0.67. The advantage of our measures is that it spans 93 countries while the standard headcount measures for 1988 are available for only 24 countries.

While the headcount index is easy to understand, it is insensitive to the degree of poverty and income transfers among the poor. Therefore, we also construct the poverty gap index that measures the extent to which individuals fall below the poverty line and the squared poverty gap ("poverty severity") index that averages the squares of the poverty gaps relative to the poverty line. Both these measures reflect the depth of poverty. While the poverty gap is insensitive to income distribution below the poverty line, the poverty severity index takes inequality among the poor into account.[8] However, the latter is not easy to interpret. As for the headcount measure, we construct the poverty gap and the squared poverty gap measures based on the two poverty line cut-offs of \$1.25 and \$2 a day. The final poverty measure we use is the Sen index given by $P_{sen} = Pov^z \left[1 - \left(1 - Gini_{poor}\right)\frac{\bar{y}_{poor}}{z}\right]$ where $\bar{y}_{poor}$ is the mean income of the poor, *Pov* is the headcount measure of poverty, $Gini_{poor}$ is the Gini coefficient among the poor and $z = 1.25, 2$ is one of two poverty lines. The Sen index captures the number of poor, the depth of their poverty, and the distribution of poverty within the poor.[9]

Poverty in the developing world is typically measured using absolute poverty lines such as \$1.25 or \$2 a day. However, since this yields zero headcount poverty rates for most developed countries, developed countries typically report poverty using relative poverty lines. These are usually expressed as a constant proportion—typically 40% to 60%—of the current

---

[8] The poverty severity index can be thought of as a weighted sum of poverty gaps, where the weights are the proportionate poverty gaps themselves.

[9] The Sen index can also be written as the average of the headcount and poverty gap measures, weighted by the Gini coefficient of the poor. In addition, the global Multidimensional Poverty Index (MPI) is a relatively new measure of acute poverty that complements traditional income-based poverty measures. Beyond income, it measures deprivations with respect to education, health and living standards. However, this measure encapsulates one of our outcome measures, schooling, so we do not use this measure. Data coverage also varies widely by year.

mean or median income (Ravallion and Chen, 2011). We construct six additional headcount measures based on relative poverty lines for all countries, developed and developing, setting the proportion at 0.4, 0.5 and 0.6 of mean and then median income. Atkinson and Bourguignon (2001) propose a hybrid version that combines absolute and relative poverty lines – absolute for low-income countries and relative for middle-income and developed countries. Their poverty line for country *i* is defined as $\max(z, k\bar{y}_i)$ where $z = 1.25, 2$ is the absolute poverty line, a reasonable lower bound for subsistence, and $\bar{y}_i$ is the mean income of country *i*. Atkinson and Bourguignon (2001) set $k = 0.37$. Based on this, we construct two Atkinson and Bourguignon hybrid poverty lines, one for $1.25 a day and one for $2 a day.[10]

2.2 Economic Development Outcomes

A rich theoretical literature links inequality to per capita income and growth of per capita income (see Galor, 2011 and Delbianco, Dabús, and Caraballo, 2014, for a summary). The classical approach emphasized the high marginal propensity to save of the wealthy and argued that rising inequality increases saving, investment, and, eventually, economic growth. In contrast, starting with Galor and Zeira (1993) demonstrated that under credit market imperfections, income distribution has a long-lasting effect on investment in human capital, aggregate income, and economic development. A second strand of literature (e.g., Alesina and Rodrik, 1994; Persson and Tabellini, 1994; Banerjee and Duflo, 2003) takes a political economy approach and shows that higher inequality leads to either greater redistribution pressures that tax physical and human capital or leads to hold up problems, with adverse consequences for economic development. Inequality can also be detrimental by encouraging socio-political instability, which is detrimental to investment (Alesina and Perotti 1996; Dutt and Mitra, 2008).

[10] See chapters in Haughton and Khandker (2009) for more details on each of these measures. Table A1 in the Appendix provides a brief description of each of the 37 measures of income distribution. The correlations between these measures range from -0.99 to 0.99. Table A2 lists the countries in our sample for which we have data on income distribution.

Inequality in the distribution of land and other natural resources can lead to under-investment in human capital, and a slower growth process (Galor and Moav, 2009). Finally, Engerman and Sokoloff (2000) and Acemoglu, Johnson, and Robinson (2005) argue that income inequality brought about oppressive institutions (e.g., restricted access to the democratic process) and simultaneously restricted access to education designed to maintain the political power of the elite and to preserve the existing inequality between the elite and the masses.

Despite the range of theoretical mechanisms that the literature alludes to, our objective is to bring new machine learning tools to bear in studying the relative importance of various measures of income distribution on development outcomes. To facilitate comparison, we follow Easterly (2007) and use the same set of economic development outcomes. The three outcome variables used by Easterly are income measured as (log) per capita income in 2002, schooling measured as secondary enrolment rates averaged over 1998-2003 (in %), and an aggregate institutional index from Kaufmann, Kraay, and Mastruzzi (2009) for 2002. We also present robustness checks with a broader array of development outcomes and measures of institutions in Appendix 2 of this paper. All outcome variables are measured at least a decade later compared to the income distribution measures.

3. A Prediction Approach

Given the sheer multiplicity of income distribution measures, we first adopt a purely predictive dimensional-reducing approach from machine learning. These approaches are well suited when we have available a large collection of possible covariates (p), possibly highly correlated as is the case here, and where the number of covariates (p) can potentially be larger than n, the number of observations. For instance, if $p = n$, an OLS estimator fits the data perfectly, but demonstrates poor out-of-sample forecasting properties because the model captures not only the signal about how predictor variables may be used to forecast the outcome, but also fits the noise present in the sample (Belloni, Chernozhukov, and Hansen, 2014a). The key question we ask is whether we can identify a parsimonious set of these multiple income distribution measures that produce the best forecast for development outcomes. We rely on traditional machine learning tools for dimension reduction based on "regularization." We impose approximate sparsity as our regularization technique – a restriction that only subset of variables *s*, where *s* << *n*, exhibit large nonzero coefficients, for the outcome variable(s) while simultaneously estimating these coefficients.

First, we apply the elastic net of Zou and Hastie (2005), which incorporates penalties from both L1 (sum of absolute value of coefficients) and L2 (sum of squared coefficients) regularization, splitting data into test (1/3) and training (2/3) sets. Elastic net is a promising approach where pairwise correlations between predictors are high, which is the case with our measures. For a linear model $y_i = \sum_{j=1}^{p} d_{i,j}\beta_j + \epsilon_i$, coefficient estimates are given by[11]

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\left(\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} d_{i,j}\beta_j\right)^2 + \lambda\left((1-\alpha)\sum_{j=1}^{p}\frac{{\beta_j}^2}{2} + \alpha\sum_{j=1}^{p}\left|\beta_j\right|\right)\right).$$

Here, λ controls the strength of the penalty, while alpha determines the relative weights of L1 and L2 regularization. When $\alpha = 0$, elastic net corresponds to ridge regression, and when $\alpha = 1$ it corresponds to LASSO (least absolute shrinkage and selection operator).[12] Following Tibshirani (1996), we choose λ using cross-validation. In cross-validation, the data are randomly divided into *K*-folds (or groups) of approximately equal sizes, with one-fold reserved for validation and the remaining *K* – 1 folds are the observations on which the model is trained. The fitted model is used to predict the responses on the test set and calculate the mean-square error.

[11] We use the model specification and the program from https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html.

[12] LASSO (alternatively lower values of α) picks a sparse model with many zero coefficients while ridge (or higher values of α) picks a more dense model with many non-zero but small coefficients.

The process is repeated K times, and the K mean-square errors are averaged and termed the cross-validation error. While λ can be chosen to minimize the cross-validation error, Tibshirani (1996) recommends using the one standard error rule when the goal is recovering the true model, instead of minimizing the prediction error. We vary α between 0 and 1 in increments of 0.1 and for each α choose the simplest (most regularized) model whose error is within one standard error of the minimal cross-validation error. Table 1 shows the variables selected for each outcome variable for three values of $\alpha = 0, 0.5, 1,$ as well as the corresponding mean square error.

Table 1: Elastic Net

Outcome variable	**Variable(s) Chosen**			**MSE on Test Set**		
	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$
Schooling	All 37 measures of income distribution	Headcount Poverty ($2 a day); Headcount Poverty ($1.25 a day); Poverty gap ($2 a day), Sen index; Atkinson-Bourguignon poverty measure at $2 a day; Decile 6	Headcount Poverty ($2 a day)	633.5	429.1	405.8
Institutional Index	All 37 measures of income distribution	Headcount Poverty ($2 a day); Deciles 1, 5, 6, 7, 9; Mehran index; Piesch	Headcount Poverty ($2 a day); Decile 1 share	0.703	0.727	0.616
Per Capita Income	All 37 measures of income distribution	Headcount Poverty, Poverty gap, Sen index, Atkinson-Bourguignon poverty measure at $2 a day	Headcount Poverty ($2 a day)	0.295	0.270	0.260

First, regardless of the outcome variable, the minimum MSE is for $\alpha = 1,$ which corresponds to the LASSO estimator. Second, as α increases from 0 to 1, fewer variables are included in the final model, but Headcount Poverty at $2 a day is always included. That gives us confidence that the inclusion of poverty at $2 a day is not accidental, despite many measures being highly correlated. Overall, elastic net picks the sparse LASSO estimator and consistently selects poverty as the income distribution measure that (negatively) impacts all three development outcomes.

Since LASSO does not yield standard errors for the variables selected, Belloni and Chernozhukov (2013) recommend using a post-LASSO procedure. For this reason, and as a robustness test, we follow their procedure, where we first apply LASSO to determine which income distribution variables can be dropped from the standpoint of predicting a particular development outcome. Subsequently, coefficients on the remaining variables are estimated via ordinary least squares regression using only the variables with nonzero estimated coefficients. Belloni and Chernozhukov (2013) show that the post-LASSO procedure works as well as and often better than LASSO in terms of rates of convergence and bias. For this procedure, we also follow Belloni, Chen, Chernozhukov, Hansen, (2012) who recommend setting the penalty level $\lambda = 2.2\sqrt{2n\log\left(\frac{2p}{0.1/\max(n,p)}\right)}$ when prediction is not the end goal. With this λ, they obtain sharp convergence results for the LASSO estimator even in the presence of heteroscedasticity.

Table 2 shows the post-LASSO procedure. Column 1 lists the outcome measure; Column 2 shows the variables picked by the LASSO procedure; Column 3 reports the post-LASSO coefficients. We find that the LASSO technique picks the headcount measure of poverty at $2 a day as the only relevant variable to predict per capita income and the institutional index, out of 37 possible measures of income distribution. For schooling, LASSO picks two poverty measures – the headcount measure and the poverty gap measure, both measured relative to the poverty line of $2 a day. However, the post-LASSO standard errors show that it is only the headcount measure that is significant at 5%. The poverty measures account for more than 50% of the variation for the income per capita and schooling outcomes, and 24% of the variation for the institutional index. Overall, from a predictive standpoint it is poverty, especially the headcount measure of poverty based on $2 a day that matters, rather than any other income distribution measure for all three development outcomes considered in Easterly (2007).

Table 2: LASSO estimator

Outcome variable	Variable(s) Chosen	Post-LASSO Coefficient	R^2	No of observations
Schooling	Poverty headcount at $2.00 a day Poverty gap at $2 a day	-0.731** (0.345) -0.332 (0.703)	0.54	80
Institutional Index	Poverty headcount at $2.00 a day	-0.015*** (0.002)	0.24	85
Per Capita Income	Poverty headcount at $2.00 a day	-0.025*** (0.002)	0.57	70

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

Since LASSO is known to be sensitive to the training sample, as a robustness check, we randomly dropped a third of the sample and reran LASSO on the remaining two-thirds of the sample. We performed this procedure 200 times. In 64% of the estimations, LASSO picks poverty when Schooling is the outcome variable; for Institutions, poverty is picked in 45% of the iterations; for per capita income, the number is 93%. This supports that poverty is a stronger predictor of per capita income and schooling as compared to institutional quality.

The results presented so far match the time span of the analyses to the one used in Easterly (2007) to facilitate comparison. As a final robustness check, we used data from the World Bank's PovcalNet to examine the sensitivity of our findings with respect to the time period chosen. The database provides information on six income distribution measures: Gini, Mean log deviation of income, Poverty headcount ($1.90 a day at 2011 PPP), poverty gap, squared poverty gap, and a Watts' poverty index.[13] To ensure that income distribution measures are lagged as much as possible relative to the outcome measures, we picked income distribution measures for the year 2010, that has measures for 90 countries. Schooling, institutional index (World Bank Governance Indicators), and per capita income (Penn World Tables) are for the year 2014 (the latest available year).

[13] This is defined as the mean across the population of the proportionate poverty gaps, as measured by the log of the ratio of the poverty line to income, where the mean is formed over the whole population, counting the nonpoor as having a zero poverty gap.

Table 3 shows the results with the same post-LASSO procedure. As with the data of earlier vintage, we obtain nearly identical results. For all three of the economic development measures, LASSO picks only the headcount measure of poverty at $1.90 a day with similar coefficient estimates, as in Table 2.

Table 3: LASSO estimator (PovcalNet data)

Outcome variable	Variable(s) Chosen	Post-LASSO Coefficient	R^2	No of observations
Schooling	Poverty headcount at $1.9 a day	-1.069*** (0.104)	0.47	77
Institutional Index	Poverty headcount at $1.9 a day	-0.021*** (0.004)	0.19	83
Per Capita Income	Poverty headcount at $1.9 a day	-0.046*** (0.004)	0.59	83

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

4. Poverty vs. Gini

The Gini coefficient, where the mean absolute difference in income is divided by mean income, measures the relative dispersion of income in the population, regardless of whether the inequality occurs at, e.g., higher or lower income levels. As a result, two income distributions with the same Gini coefficient (and the same mean income) can have different poverty levels, with one being clearly preferred to another by a policymaker. The importance of assessing the entire distribution (as opposed to a few summary measures, which would be sufficient if the shape of the distribution were fixed, e.g., lognormal) is well known in decision theory, and a similar logic applies to comparing income distributions, as illustrated below using an example from Menezes, Geiss, and Tressler (1980).

Consider Country 1, with 50% of the population earning $1 per day, and 50% earning $2 per day. The Gini coefficient for this country is 1/6. Country 2 started with the same income distribution as Country 1, but then went through some government interventions that changed the income distribution of the poorer part of its population (without changing the mean income)

– so that those who were earning $1 per day split in two equal groups, earning either $0 or $2 per day. Thus, in Country 2, 25% of the population earn $0 per day and 75% earn $2 per day. Gini coefficient for Country 2 is ½, greater than that for Country 1. Finally, consider Country 3 that also started with the same income distribution as Country 1, but where income distribution of the wealthier part of the population was changed – those earning $2 per day split in two equal groups, earning either $1 or $3 per day – so in Country 3, 75% of the population earn $1 per day and 25% earn $3 per day. Gini coefficient for Country 3 is ½ – the same as for Country 2. In the terminology of Menezes et al. (1980), income distributions in Country 2 and Country 3 differ by a mean-variance preserving transformation. Compared to Country 1, Country 2 has more risk in a lower tail of the income distribution, and Country 3 has more risk in its upper tail.

Which of these three income distributions is better from a poverty perspective? Though Country 2 and 3 have a greater Gini coefficient than Country 1, the proportion of population strictly below the poverty line is not necessarily higher in these countries, as Table 4 illustrates.

Table 4: Gini and Poverty Rankings

Measures	**Proportion of Population below Poverty Line**		
	Country 1	*Country 2*	*Country 3*
Poverty at $1.0 per day	0%	25%	0%
Poverty at $1.25 per day	50%	25%	75%
Poverty at $2.5 per day	100%	100%	75%
Gini coefficient	$^1/_6$	½	½

With a poverty measure of $1.0 per day, Country 2 is the worst; at $1.25 per day, Country 3 is the worst; and at $2.5 per day, Country 3 is the best. This is because increasing inequality for the part of the population that is below the poverty line decreases poverty, while increasing inequality for the part of the population above the poverty line increases poverty. More broadly, an outward shift in the Lorenz curve, indicating a rise in the Gini coefficient while holding the mean income constant, can be consistent with either an increase or decrease in the widely used headcount measure of poverty.

If income is distributed as log-normal, then two parameters are sufficient to summarize the entire income distribution. For instance, for a log-normal distribution, poverty can be written as a non-linear function of mean income and the Gini coefficient. However, as Battistin, Blundell and Lewbell (2009) show using detailed data from US households, there are significant departures from log normality in the income data. They find that the log of income is far from normal with upper tail skewness and greater kurtosis. In contrast, the log of consumption is very close to normal for US households. They attribute this difference to a larger transitory component in income, a component that would be arguably more pronounced in developing countries that form a large part of our sample. Cowell (2011) shows that while the lognormal distribution holds for particular segments of relatively homogeneous groups within a country, it breaks down for the income distribution of the aggregate population of a country. Therefore, systematic departures from lognormality are evident and expected in many earnings distribution. Finally, even if some income distribution measures can be written as a function of others, introducing high collinearity, such high correlations are not problematic for LASSO prediction (Hebiri and Lederer, 2013).

Given that the results in Section 3 identify poverty as an important predictor of development outcome and given that the literature has overwhelmingly focused on the Gini coefficient, we next evaluate the relative importance of poverty vs. inequality (as measured by the Gini coefficient) for economic development.

5. Causal Inference for Poverty vs. Inequality

Our LASSO approach consistently identified poverty as having a strong association to each of the three development outcomes within a sparse framework. However, such a predictive procedure does not allow us to draw inferences about model parameters since model selection mistakes cannot be ruled out (Belloni, Chernozhukov, and Hansen, 2014a). In particular, LASSO targets prediction rather than estimating specific parameters or coefficients of interest (the coefficient on poverty and/or Gini in our case). LASSO also drops certain variables that have small but non-zero coefficients, but then we have the standard omitted variable bias problem, making inferences problematic. Therefore, for inferring the relative importance of poverty vs. inequality, we adopt the "double selection" procedure (Belloni, Chernozhukov, and Hansen, 2014b), that allows for valid inferences even in the presence of selection mistakes.

Let y_i be a particular development outcome measure (schooling, institutional quality, per capita income), d_i be the income distribution measure(s) of interest (headcount poverty and/or Gini) whose impact we would like to infer, and X_i be a vector of controls. The standard approach is to estimate a linear model:

$$y_i = \beta d_i + X_i \Theta_y + \xi_i \qquad (1)$$

where ξ_i is the error term and the objective is to conduct inference on β.

For valid inferences, the key identifying assumption is that d_i may be taken as randomly assigned once a sufficient set of factors in X_i have been controlled for. This is a strong assumption and estimating a structural effect when relying on such a "conditional on observables" argument requires knowing which controls to include. Otherwise, we run the risk of a key omitted variable driving both distribution and development outcome. Some ways to do this is to rely on past work (e.g., conditioning on the controls used in Easterly, 2007) or on economic intuition, or on theory to explicitly define what variables belong in the regression. While the first two options are arguably ad-hoc, even theory may not help narrow the multiplicity of regressors – theories are not mutually exclusive, and different models can identify different variables that truly belong in (1). Therefore, we have a potentially vast set of controls to choose from, and in a cross-sectional context, the number of regressors may easily exceed *n*, the number of observations. Even if we rely on theory to identify a small number of controls that enter the model of interest, we can rarely say with confidence that a linear functional form (as assumed in equation 1) is appropriate. Again, we are left with various transformations and interactions of a set of variables, which may again lead to the number of regressors exceeding the number of observations. In both cases, choosing appropriate controls and functional forms is essentially a dimensionality problem and is well recognized in economic development (see Leamer, 1985, Sala-i-Martin et al., 2004; Levine and Renelt, 1992; Sala-i-Martin, 1997).

Since some structure needs to be necessarily imposed, high-dimensional sparse models assume exogeneity of d_i once we control linearly for a relatively small number of variables in X_i whose identities are a priori unknown. This assumption, termed approximate sparsity, implies that a linear combination of these unknown controls produces relatively small approximation errors and allows us to approach the problem of estimating β as a variable selection problem. Standard machine learning procedures are used to reduce the number of variables to a manageable size. With approximate sparsity, equation (1) now includes an approximation error term r_{yi} in the outcome equation.[14]

$$y_i = \beta d_i + X_i \Theta_y + r_{yi} + \xi_i \tag{2}$$

where $E[\xi_i | d_i \, X_i \, r_{yi}] = 0$

The “double-selection” procedure Belloni, Chernozhukov, and Hansen (2014b) has three steps. In the first step, an additional reduced form relation between the treatment and controls is introduced

$$d_i = X_i \Theta_d + r_{di} + v_i \tag{3}$$

where $E[v_i | \, X_i \, r_{di}] = 0$. This step selects control variables that are strongly related to the

[14] It is assumed that r_{yi} is small enough relative to sampling error. See Belloni, Chen, Chernozhukov, and Hansen (2012).

variable of interest d_i and thus potential confounders ensuring the validity of post-model-selection inference. In the second step, we substitute (3) in (2) and estimate a reduced form for y_i,

$$y_i = X_i[\Theta_y + \beta\Theta_d] + r_{yi} + \beta r_{di} + (\xi_i + \beta v_i). \qquad (4)$$

Equations (3) and (4) are predictive relationships, which may be estimated using high-dimensional methods. The two in conjunction help guard against omitted variable bias. Applying variable selection to (4) keeps the residual variance small, and helps identify important confounds, guarding against omitted-variable bias. Similarly, applying variable selection to (3) assures that we include controls that have strong predictive power for d_i. Even if these are only moderately correlated with the outcome variable, not including them may inappropriately attribute the effect to d_i, biasing inference. In the final step, we estimate β, the effect of interest, by a linear regression of y_i on d_i and the union of the set of variables selected in the first two variable selection steps.

To permit full comparison, we use the 1988 headcount poverty measure based on $2 a day poverty line and Easterly's measure of inequality, which is the Gini coefficient derived by adjusting data from the WIDER (2000). Easterly uses a Gini coefficient which is averaged over the time period 1960-1998. For the full set of control variables, we use 67 variables from Sala-i-Martin et al (2004) all measured at or close to the year 1960.[15] To ensure that poverty is not simply capturing low income in 1988 and hence higher potential for growth, we always include the growth rate of per capita GDP between 1988 and 2002 as a control. Finally, we add two dummies for the legal origins of countries (British and French), since these are used in Easterly (2007). All outcome measures are from the year 2002 so they are not mechanically correlated with any of the independent variables. Please see Table A3 in the appendix for more details on these variables and Table 1 in Sala-i-Martin et al. (2004) for original data sources.

[15] 1988 is the earliest year for which poverty measures for a sufficiently large set of countries are available.

5.1 Results From Double-Selection Procedure

Tables 5A-5C present our cross-sectional results for each of the three outcome variables, namely schooling, institutions, and income. Column 1 in each table uses OLS to replicate the Easterly findings - inequality is associated with a lower level of schooling, poorer institutional quality, and a lower level of per capita income. Column 2 continues to use OLS but adds poverty to the Gini coefficient. When we use the two distribution measures in conjunction, we find that poverty matters strongly for schooling, institutional quality, and the level of development. However, the Gini coefficient is no longer significant for per capita income and matters only weakly for schooling. The coefficient estimate declines sharply and is only marginally significant (p-value = 0.099).

In Column 3, in each of the three tables, we implement the double-selection procedure where $d_i = Gini$ and where we subsume poverty into the vector of controls X_i. That is, here inference is solely for inequality. We find that the LASSO procedure always selects poverty as for each of the outcome measures in the reduced-form equation (3), consistent with our findings in Tables 1 & 2, where we attempted to predict the outcome variable. Running OLS of the outcome measures on Gini and the union of controls selected in equations (2) and (3) shows that the Gini coefficient is either insignificant (for schooling) or has the wrong sign (for institutions and per capita income). In contrast, for all three measures, poverty matters strongly. This suggests that poverty is an important omitted variable in regressions that evaluate how inequality matters for development outcomes, and its inclusion renders inequality insignificant.

In Column 4 we do the reverse – we subsume the Gini coefficient in the vector of controls X_i and set $d_i = Poverty$. Across outcome measures, we find that the LASSO procedure does not choose Gini, so inequality does not seem to be an important omitted variable for either predicting poverty or the outcome measure. At the same time, the double-selection procedure shows that poverty matters for schooling and per capita income, but not institutional quality.

Finally, Column 5 in all three tables sets d_i = {*Gini, Poverty*}, estimates three reduced-form equations, one each for inequality and poverty and one for the outcome measure, and then presents OLS estimates for each outcome measure on Gini, poverty, and the union of controls selected. In this more demanding specification, poverty matters for all three development outcome measures, while Gini does not. Overall, these results suggest that the percentage of people below internationally comparable poverty lines is more important than the oft-used single statistic, Gini coefficient.

To get a sense of the magnitude of the effect of poverty on development outcomes, compare Uruguay, at the 25th percentile of the poverty measure in our sample, with 10% below the poverty line, to Bolivia at the 75th percentile, with 50% of the population below the poverty line. Our estimates imply that reducing poverty in Bolivia to the level of Uruguay leads to an increase in secondary enrolment rates by 20%, which is nearly identical to the difference in secondary enrolment rates of the two countries. Similarly, Bolivia's institutional index would improve by 0.2, equivalent to a move in Bolivia's institutional rank of 48 out of 63 countries to a rank of 36. Uruguay's actual institution rank is 26, so poverty explains only part of the gap in institutions. Finally, such a decline in poverty would increase Bolivia's per capita GDP by 0.2 log points, approximately a 22% increase. In reality, there is a 1.07 log point difference between the per capita GDP of Uruguay and Bolivia, so while the effect is substantive, it is far less than the actual income gap between Uruguay and Bolivia. All comparisons are based on the estimated coefficients in Column 5 of Tables 5A-5C.

Table 5A: Inequality, Poverty, and Schooling

	(1)	(2)	(3)	(4)	(5)
	OLS	*OLS*	*Double-selection*	*Double-selection*	*Double-selection*
Gini	-1.474***	-0.486*	-0.025		-0.109
	(0.292)	(0.291)	(0.344)		(0.404)
Poverty at $2.00 a day		-0.820***	-0.298***	-0.422***	-0.509***
		(0.089)	(0.066)	(0.146)	(0.167)
Growth of per capita GDP (1988-2002)			-0.766	-0.296	-1.042
			(0.839)	(1.227)	(1.320)
Absolute Latitude			0.283	0.297	0.277
			(0.201)	(0.248)	(0.231)
Population Growth Rate 1960-90			-862.845		-610.629
			(662.126)		(568.215)
Fertility in 1960s			-9.417		
			(15.365)		
Life Expectancy in 1960			0.370	0.116	-0.215
			(0.390)	(0.559)	(0.523)
African Dummy			-38.975***		
			(8.004)		
Latin American Dummy			-8.471		
			(5.319)		
Fraction Population Less than 15			-42.903		
			(84.046)		
Primary Exports 1970			30.600***		
			(11.370)		
Malaria Prevalence in 1960s				-16.581**	-15.774*
				(7.547)	(8.398)
Primary Schooling in 1960				38.333	46.573**
				(22.950)	(22.313)
Years Open 1950-94				-0.948	-2.966
				(8.127)	(8.386)
Air Distance to Big Cities				0.001	0.001
				(0.001)	(0.002)
Colony Dummy				-6.416	1.108
				(6.425)	(7.716)
Gov. Consumption Share 1960s				42.090	40.617
				(61.771)	(63.431)
Fraction Population Over 65				236.325	-26.806
				(183.054)	(252.334)
Interior Density					0.037
					(0.029)
European Dummy					12.019
					(13.723)
Observations	120	82	60	59	59
R-squared	0.15	0.56	0.85	0.80	0.82
Joint significance test	25.47***	60.48***	78.44***	35.00***	34.04***

The outcome variable is secondary enrolment rates averaged over 1998-2003; Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%; All columns include a constant (not shown)

Table 5B: Inequality, Poverty, and Institutions

	(1)	(2)	(3)	(4)	(5)
	OLS	*OLS*	*Double selection*	*Double selection*	*Double selection*
Gini	-0.031***	-0.023**	0.013*		0.009
	(0.006)	(0.009)	(0.007)		(0.006)
Poverty at $2.00 a day		-0.012***	-0.005**	-0.004	-0.005*
		(0.002)	(0.002)	(0.003)	(0.003)
Growth of per capita GDP (1988-2002)			0.043	0.047*	0.045
			(0.033)	(0.028)	(0.030)
European Dummy			0.406*	0.490**	0.528**
			(0.208)	(0.226)	(0.241)
Fertility in 1960s			-0.983***	-0.745**	-0.712*
			(0.299)	(0.318)	(0.383)
Fraction Protestants			0.275**	0.298**	0.311**
			(0.130)	(0.122)	(0.131)
Fraction Population In Tropics				-0.382**	-0.345*
				(0.164)	(0.181)
Years Open 1950-94				0.386**	0.350*
				(0.187)	(0.196)
Air distance to Big Cities				0.051	0.038
				(0.052)	(0.054)
Colony Dummy				0.025	0.024
				(0.112)	(0.116)
Population Density (Coastal) in 1960s				0.000*	0.000**
				(0.000)	(0.000)
Gov. Consumption Share 1960s				1.233	1.598
				(1.056)	(1.021)
Life Expectancy in 1960				-0.004	-0.006
				(0.010)	(0.011)
Fraction Orthodox				-0.606***	-0.608***
				(0.146)	(0.162)
Fraction Population Over 65				-0.781	-1.508
				(2.966)	(3.351)
Population Growth Rate 1960-90			-14.806		-6.041
			(13.289)		(12.440)
Latin American Dummy			-0.184		
			(0.140)		
Fraction Population Less than 15			2.833		
			(1.900)		
Primary Exports 1970			-0.141		
			(0.236)		
Observations	128	87	63	63	63
R-squared	0.13	0.29	0.85	0.89	0.89
Joint significance test	24.24***	25.73***	55.73***	44.41***	39.85***

The outcome variable is an aggregate institutional index from Kaufmann, Kraay, and Mastruzzi (2009) for 2002; Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%; All columns include a constant (not shown)

Table 5C: Inequality, Poverty, and Per Capita Income

	(1)	(2)	(3)	(4)
	OLS	*Double selection*	*Double selection*	*Double selection*
Gini	-0.009	0.004		0.005
	(0.009)	(0.005)		(0.006)
Poverty at $2.00 a day	-0.024***	-0.006***	-0.005**	-0.006**
	(0.002)	(0.002)	(0.003)	(0.002)
Growth of per capita GDP (1988-2002)		0.027	0.008	0.019
		(0.019)	(0.018)	(0.020)
GDP per capita in 1960 (log)		0.606***	0.643***	0.665***
		(0.058)	(0.068)	(0.075)
Political Rights			0.012	
			(0.034)	
European Dummy			0.160	0.158
			(0.140)	(0.124)
Fertility in 1960s		-0.268	-0.046	0.020
		(0.203)	(0.164)	(0.189)
Life Expectancy in 1960		0.015**	0.017	0.012
		(0.007)	(0.010)	(0.010)
Primary Schooling in 1960			0.079	0.287
			(0.333)	(0.346)
Fraction Population In Tropics			-0.224	
			(0.135)	
Years Open 1950-94		0.291***	0.313**	0.219*
		(0.098)	(0.123)	(0.112)
Air distance to Big Cities			0.005	0.009
			(0.026)	(0.034)
Colony Dummy			-0.110	-0.092
			(0.084)	(0.081)
Gov. Consumption Share 1960s			-0.725	-0.354
			(0.683)	(0.685)
Fraction Hindus			0.288	
			(0.229)	
Fraction Population Over 65			-4.341*	-4.066*
			(2.220)	(2.097)
Latin American Dummy		-0.046		
		(0.073)		
Fraction Population Less than 15		0.646		
		(0.895)		
Absolute Latitude				0.004
				(0.003)
Primary Exports 1970				-0.087
				(0.227)
Population Density (Coastal) in 1960s				0.000
				(0.000)
Fraction Orthodox				0.038
				(0.128)
Observations	72	64	61	60
R-squared	0.57	0.96	0.96	0.97
Joint significance test	72.63***	270.1***	181.6***	137.4***

The outcome variable is per capita GDP in 2002; Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%; All columns include a constant (not shown)

In terms of the controls chosen by LASSO, we find that for per capita income, the controls chosen are the log of per capita GDP in 1960 (measuring initial mean income), openness measured by the Sachs-Warner index, and the fraction of population above 65 years. For institutions, the controls selected by LASSO and ones that are significant are the fraction of population living in the tropics, coastal population density, the Sachs-Warner openness measure, a dummy for European countries, and the fraction of Protestants and Orthodox. Finally, for schooling, the variables that matter significantly are primary education in 1960 and malaria prevalence in 1960s.[16]

5.2 Results from Instrumental Variables and Double-Selection Procedure

Estimating the impact of poverty on development outcomes results in an endogeneity problem. While the previous variable selection methodology systematically attempts to address the omitted variable bias, our choice of the Sala-i-Martin variables as controls is necessarily incomplete. Additionally, even though we measure the income distribution in the year 1988, we are not wholly immunized to the reverse causality issue. For example, schooling and institutions are slow to adjust and may impact poverty and inequality. Easterly (2007) takes this endogeneity issue seriously and attempts to resolve it with a creative new instrument based on land endowments drawing upon the work of Engermann and Sokoloff (1997).

Engermann and Sokoloff (1997) argue that land endowments are a central determinant of inequality. Land endowments, for instance, in Latin America, were suitable for the cultivation of commodities such as sugar at large scale and the use of slave labor, which was, in

[16] We also checked the relative importance of poverty vs. Gini for seven additional development outcomes. For life expectancy, infant mortality, and maternal mortality, we find that poverty is strongly associated with these outcomes (negative for life expectancy and positive for the mortality measures). The Gini coefficient either does not matter or has the wrong sign. For a happiness index (from the latest World Happiness Report) neither income distribution measure matters. Poverty (but not Gini) also matters for primary education (parallel to our findings for secondary education). Finally, for a measure of trust from the World Values Survey, it is inequality rather than poverty which is detrimental for trust in others. These results are presented in Table A4 in Appendix 2.

turn, associated with high inequality and even poverty. In North America, the endowments led to wheat cultivation, smaller-scale family farms, encouraging the growth of a middles class, and lower inequality. High levels of inequality, in turn, have a deleterious impact on the quality of institutions, the level of human capital investment, and, ultimately economic development. Therefore, as an instrument, Easterly (2007) uses the suitability of arable land for wheat vs. sugarcane. The justification is that land endowments are plausibly exogenous. More precisely, the key exclusion restriction for the instrumental variable is that current per capita incomes, schooling, and institutions, even if persistent, are unlikely to be strongly influenced by land endowments except through one channel, which is inequality.

We use the wheat-sugar ratio defined as

$$\ln \frac{(1 + share\ of\ arable\ land\ for\ wheat)}{(1 + share\ of\ arable\ land\ for\ sugar)}$$

as the key instrument (ES instrument). We take as given that this is a valid instrument but examine if the instrument affects development outcomes through the inequality channel or the poverty channel. In regressions with either poverty or Gini as the only independent variable, we use only the wheat-sugar ratio as an instrument. In results that instrument both Gini and poverty, we use the share of the country's cultivated land area in tropical climate zones from Sachs and Warner (1997) as a second instrument.[17]

To facilitate comparison with Easterly (2007) we first instrument Gini and poverty one at a time, while including the other measure as an uninstrumented independent variable (see Columns A and B in Table 6). Subsequently, we instrument both measures with the ES instrument and the share of the country's cultivated land area in tropical climate zones in Columns C of Table 6. Table 6 also reports the 1st-stage F-statistic to evaluate the strength of

[17] Easterly (2007) uses this as a second instrument to conduct overidentification tests.

the instrument.[18] Columns 1A, 2A, and 3A show that when we instrument Gini but include poverty as an additional control, the inequality results of Easterly weaken considerably.[19] Inequality is insignificant for per capita income but matters for institutional quality and schooling. When we instrument only for poverty in Columns 1B, 2B, and 3B, we find that poverty is significant for all three outcome variables. Now, Gini does not matter at all. The results in Columns A mean that even if the ES instrument is plausibly exogenous and not subject to the weak-instrument critique, the exclusion restriction assumption in Easterly (2007) is questionable. The results in columns B imply that the instrument works better for poverty and that the effect of poverty on development outcomes is relatively more robust to the inclusion of Gini. Land endowments seem to affect development outcomes more strongly through its impact on poverty rather than the Gini coefficient. When we use the two in conjunction, in Columns 1C, 2C, and 3C, we find that it is only poverty that matters for per capita income and schooling, while neither distributional measure matters for the institutional index.

[18] In all cases, the first-stage F-statistics are well above the critical values from Stock and Yogo (2004) so that the ES instrument is not subject to the weak instrument critique. We are unable to test for over-identification restrictions since our system is just-identified.

[19] Easterly interprets the increase in coefficient on inequality for the IV results as an underestimation of the causal relationship by the OLS specification. However, it may also be interpreted as attenuation due to measurement error in the inequality measure, which Easterly acknowledges when discussing the data sources for inequality.

Table 6: Inequality, Poverty and Development Outcomes (IV results)

	(1A)	(1B)	(1C)	(2A)	(2B)	(2C)	(3A)	(3B)	(3C)
	Secondary School Enrolment			Institution Index			Per Capita Income (log)		
Gini	-1.175*	-0.335	0.408	-0.058**	-0.010	-0.050	-0.031	-0.002	0.059
	(0.713)	(0.401)	(1.745)	(0.027)	(0.015)	(0.059)	(0.024)	(0.017)	(0.103)
Poverty at $2.00 a day	-0.765***	-0.986***	-1.182**	-0.007*	-0.020***	-0.009	-0.021***	-0.029***	-0.045*
	(0.108)	(0.196)	(0.483)	(0.004)	(0.007)	(0.014)	(0.003)	(0.008)	(0.026)
Observations	78	78	78	82	82	82	67	67	67
Joint significance test	54.16***	30.91***	25.17***	15.44***	14.89***	13.28***	55.21***	23.94***	13.49***
1st stage F-statistic for Gini	15.86***		16.28	15.30***		16.91***	12.25		14.28***
1st stage F-statistic for poverty		17.63***	23.89***		18.57***	26.51***		15.07***	23.11***

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%; All columns include a constant (not shown).
Columns 1A, 2A, and 3A instrument only inequality; Columns 1B, 2B, and 3B instrument only poverty; Columns 1C, 2C, and 3C instrument both poverty and inequality.

Table 7: Inequality, Poverty and Development Outcomes (IV + Double-Selection results)

	(1A)	(1B)	(1C)	(2A)	(2B)	(2C)	(3A)	(3B)	(3C)
	Secondary School Enrolment			Institution Index			Per Capita Income (log)		
Gini	0.529		0.370	0.050		0.035	-0.022		-0.003
	(1.336)		(1.149)	(0.062)		(0.028)	(0.035)		(0.016)
Poverty at $2.00 a day	-0.403***	-0.576***	-0.650***	-0.006	0.007	0.001	-0.004	-0.008***	-0.007**
	(0.084)	(0.218)	(0.213)	(0.004)	(0.006)	(0.008)	(0.004)	(0.003)	(0.003)
Growth of per capita GDP	0.326	0.493	0.822	0.072**	0.076**	0.109***	-0.002	0.006	0.007
(1988-2002)	(1.264)	(1.244)	(1.747)	(0.035)	(0.032)	(0.038)	(0.022)	(0.020)	(0.026)
Observations	58	57	57	60	59	59	59	57	57
Joint significance test	39.44***	26.29***	21.56***	36.99***	30.87***	31.7***	137.01***	177.80***	57.28***
1st stage F-statistic for Gini	0.88		1.39	0.40		2.17	1.06		1.36
1st stage F-statistic for Poverty		4.95***	5.02***		3.97**	4.36**		7.32***	6.07***
OID test p-value	0.39	0.77	0.47	0.22	0.38	0.33	0.59	0.94	0.72

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%; All columns include a constant & controls chosen by LASSO (not shown). Columns 1A, 2A, and 3A instrument only inequality; Columns 1B, 2B, and 3B instrument only poverty; Columns 1C, 2C, and 3C instrument both poverty and inequality.

Easterly (2007) rules out various competing hypotheses by including controls, one at a time, for ethnolinguistic fractionalization, legal origin and tropical land.[20] Essentially, fractionalization, legal origin, and tropical land are also plausibly exogenous and persistent. Therefore, if the inequality hypothesis is indeed correct in explaining development outcomes, then it is important to control for these variables. Controlling for these variables also implies that identification relies on variation in the wheat-sugar land endowment ratio that is not related to any of these variables. But as argued in the previous section, the set of controls remain necessarily incomplete. One can easily argue that colonial origin, demographic composition, disease burden, religious composition, etc. are also equally persistent, and not controlling for these channels may invalidate the exclusion restriction. Therefore, to evaluate the robustness of our findings in Table 6, we combine the Easterly instruments with the double-selection procedure in Chernozhukov, Hansen, and Spindler (2015).

The double-selection procedure uses Lasso-based methods to again identify the key set of controls and then performs inference using IV estimation. With a single instrument and a single endogenous variable, we now have a three-equation system

$$y_i = \beta d_i + \mathrm{X}_i \Theta_y + \xi_i \tag{5}$$

$$d_i = \gamma I_i + \mathrm{X}_i \Theta_d + v_i \tag{6}$$

$$I_i = \mathrm{X}_i \Theta_I + u_i . \tag{7}$$

The first relates the development outcome y_i (one of schooling, institutions or income) to a measure of income distribution d_i,(poverty or inequality) and vector of controls, the second relates the income distribution measure to an instrument (the wheat-sugar ratio or tropical land) and controls, while the third relates the instrument I_i to the control variables. We can again write this as three reduced-form equations relating the structural variables to the controls

[20] Easterly (2007) uses share of tropical land subsequently as an additional instrument since previous work has demonstrated that in tropical countries, rich elites adopt extractive strategies that were designed to capture rents for the elites (Acemoglu et al., 2005).

$$y_i = \mathrm{X}_i\widetilde{\Theta}_y + \tilde{\xi}_i \tag{5'}$$

$$d_i = \mathrm{X}_i\widetilde{\Theta}_d + \tilde{v}_i \tag{6'}$$

$$I_i = \mathrm{X}_i\widetilde{\Theta}_I + \tilde{u}_i. \tag{7'}$$

As in the previous section, we select a set of control terms from the variables in Sala-i-Martin et al. (2004) with a LASSO variable selection procedure. Valid estimation and inference for the parameter of interest β proceeds by conventional IV method of using the instrument(s) for income distribution and the union of variables selected from each reduced form as included controls. This is easily extended to a scenario where we have more than one endogenous variable and more than one instrument. For example, with two endogenous variables (poverty headcount and Gini) and three instruments (needed for overidentification tests with two endogenous variables), we would apply LASSO to 6 equations.

In Table 7, in Column 1A, 2A and 3A, we treat Gini as the sole endogenous variable and use the wheat-sugar ratio (logged), the share of the country's cultivated land area in tropical climate zones and fraction of land in tropical climate zone as instruments.[21] As before, we subsume poverty into the vector of controls and use LASSO to pick from this set. In all three columns, we see that LASSO picks the headcount measure of poverty. Table 7 includes but does not report the union of controls chosen from equations (5')-(7'). At the same time, the Gini coefficient is no longer significant for all three of the development outcomes – a finding very similar to the simpler specification in Table 6. At the same time, we should be cautious since the first-stage F-statistic shows that the instruments are weak. Essentially, the LASSO procedure picks dummies for Latin America, sub-Saharan Africa, and the fraction of population younger than 15, all of which have a strong positive relationship with Gini in the first-stage.[22]

[21] The fraction of land in tropical climate zone is based on the Hodridge life zones system – a global bioclimatic scheme for the classification of land areas. The variable is from Sala-i-Martin (2004) and is plausibly more exogenous. It is related to but distinct from the share of cultivated land in the tropical climate zone.

[22] If we use the wheat-sugar ratio as the sole instrument, then the coefficient on Gini is positive and insignificant for all outcome measures in Table 7.

Columns 1B, 2B, and 3B treat poverty as endogenous and subsume Gini in the vector of controls for LASSO variable selection. LASSO does not select the Gini coefficient in either the outcome equation, the equation for poverty, or for any of the three instruments. Here we find that poverty matters negatively for schooling and per capita income, but not for the institutional index. The model in Columns 1C, 2C, and 3C treats Gini and poverty as potentially endogenous, performs LASSO on 6 equations (one for the outcome measure, two measures of income distribution and three instruments), and then runs OLS of each development outcome on Gini and poverty with the union of controls chosen by LASSO. Results are similar to those in Column B. Once again, poverty is the variable that matters for development outcomes. Comparing the estimates to the post-LASSO OLS estimates in Table 3-5 (last column) we see that the estimated coefficients on poverty for both schooling and per capita income increase substantially, suggesting the presence of attenuation bias in the OLS estimates. Not surprisingly, since Table 7 includes multiple controls, the magnitude of the coefficient estimates declines substantially compared to Table 6. For both income distribution measures, across specifications, the Hansen J-test fails to reject the overidentification restrictions.

At the same time, these results should be interpreted with caution. The inclusion of multiple controls also renders the instruments weak – while the first-stage F-statistic is significant, it is small and does not exceed the critical Stock-Yogo critical values. With weak instruments, the 2SLS estimates are biased towards the OLS estimates, the tests of significance have incorrect size, and confidence intervals are wrong. Therefore, first, we compared these 2SLS estimates with LIML estimates (limited information maximum likelihood). LIML estimates have better small sample properties than 2SLS with weak instruments – they are a weighted combination of the OLS and 2SLS estimate, and the weights happen to be such that they (approximately) eliminate the 2SLS bias. We find that both the coefficients and standard error for the LIML estimates are very close to the 2SLS estimate which is somewhat reassuring.

Next, we substituted equation (7) into (6) and estimated a reduced-form regression of dependent variables on the three instruments. Testing that the coefficients on the instruments equal 0 tests the hypothesis that $\beta = 0$. This procedure is robust to weak instruments since no information about the correlation between suspected endogenous income distribution measures and instruments are required.[23] With no controls, all three instruments are individually and jointly significant for each of the three outcome measures. Once we add the LASSO based controls, things do change. We find that for schooling and per capita income, the log of the wheat sugar ratio and the fraction of land in tropical climate zone are individually significant, and all three instruments are jointly significant. In contrast, with institutions as the dependent variable, the three instruments are neither individually nor jointly significant. That is, once we add appropriate controls chosen by the LASSO selection procedure, we run into the weak instrument problem for institutions.

The results shown in Table 7 is the most demanding specification, combining instruments with an exhaustive set of controls, chosen by LASSO. A generous interpretation of these findings is that when it comes to economic development, it is poverty rather than inequality that plays a more important role. At the least, it should raise questions whether even in a cross-sectional setting, inequality has a robust and negative impact on development outcomes. A more realistic assessment is that even with plausibly exogenous instruments, the exclusion restriction for the ES instrument is questionable – that it works through poverty rather than inequality. Moreover, even the impact of poverty and the magnitude of this impact are sensitive to the use of appropriate set of controls, and to the exact development outcome of interest. Our results with respect to schooling and per capita income are more or less consistent across all specifications and methodologies used in the paper. However, for the institutional

[23] Angrist and Krueger (2001) advocate this approach to testing the null hypothesis that coefficient of interest equals zero when one is worried about weak-identification.

index, the most demanding specification shows a negligible role for income distribution in general, regardless of whether we use poverty or inequality to capture distribution. Overall, our results indicate that it is inequality at the bottom, as measured by poverty that matters more for development outcomes than inequality at the top.

6. Conclusion

It is widely acknowledged that Gini coefficient does not tell the whole story about income distribution – in particular, it is equally affected by increase in inequality at the top and at the bottom. At the same time, despite a plethora of measures of income distribution, most of the previous research has almost exclusively relied on the Gini coefficient as the measure of income inequality. We draw upon recent advances in causal inferences with high-dimensional sparse models that combine insights from machine learning with causal inferences. Within a predictive framework, we first show that LASSO chooses only the headcount poverty measure based on a poverty line of $2 a day as the key income distribution statistic in predicting, schooling, institutional quality, and per capita income, among a whole host of income distribution measures. Next, we employ post-LASSO models for causal inference that help guard against omitted variable bias and model selection mistakes. We show that the fraction of population living in poverty matters much more strongly for development outcomes than does the Gini coefficient, with LASSO chosen controls accounting for a host of confounding influences. Finally, instrumental variable estimates in conjunction with post-LASSO models show that compared to the Gini coefficient, poverty is more strongly associated with schooling and per capita income. At the very least, our results imply that the causal link from inequality (as measured by Gini) to development outcomes is tenuous.

Our finding calls for a focus on inequality at the bottom – that tackling poverty may yield additional dividends in terms of increasing income, improving schooling, and institutional quality. The set of policies and instruments most suited to tackling poverty are different from ones that address the challenge of income and wealth being increasingly concentrated amongst the top 1% or the top 0.1%.

We are aware of the challenges of establishing causality in such cross-country research. Using recent techniques developed in the field of machine learning and extended to a framework for causal inference allows us to account for all sorts of confounding influences systematically. At the same time, we acknowledge that these do not meet the gold standard of a randomized trial or a perfect natural experiment. Future research, relying on more in-depth and detailed datasets, and more plausible identification mechanisms, may help address the causal links between income distributions and development outcomes more convincingly. Future research can also productively explore if the direction of causal links between income distribution and economic development varies with the level of development. For instance, in developing countries, it is plausible that income distribution affects development outcomes while in advanced economies, investments in schooling and institutional quality constraint the rise in income inequality.

References

Acemoglu, Daron, Johnson, Simon, Robinson, James. 2005. Institutions as the fundamental determinant of long term growth. In: Aghion, Philippe, Durlauf, Steven (Eds.), Handbook of Economic Growth. Elsevier.

Alesina, A. and Perotti, R. (1996). Income Distribution, Political Instability, and Investment, European Economic Review 40(6): 1203—1228.

Alesina, Alberto, Rodrik, Dani, 1994. Distributive politics and economic growth. Quarterly Journal of Economics 108, 465–490.

Angrist, J.D., Krueger, Alan B., 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. Journal of Economic Perspectives 15, 69–85.

Atkinson, Anthony. 1987. On the measurement of poverty, Econometrica, 55: 749-764.

Atkinson, Anthony B., and Francois Bourguignon, 2001. Poverty and Inclusion from a World Perspective,'' in Joseph Stiglitz and Pierre-Alain Muet (Eds.), Governance, Equity and Global Markets (New York: Oxford University Press).

Banerjee, Abhijit V., Duflo, Esther, 2003. Inequality and growth: what can the data say? Journal of Economic Growth 8, 267–299.

Banerjee, A.V. and Mullainathan, S., 2008. Limited attention and income distribution. The American Economic Review, 98(2), 489-493.

Barro, R.J., 2000. Inequality and Growth in a Panel of Countries. Journal of Economic Growth 5, 5–32.

Battistin, E., Blundell, R. and Lewbel, A., 2009. Why is consumption more log normal than income? Gibrat's law revisited. Journal of Political Economy, 117(6), 1140-115.

Beck, Thorsten, Demirgüç-Kunt, Asli, Levine, Ross, 2007. Finance, inequality and the poor. Journal of Economic Growth 12, 27-49.

Belloni, A. and Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli, 19(2), 521-547.

Belloni, Alexandre, Chernozhukov, Victor and Hansen, Christian B., 2013. High-Dimensional Sparse Econometric Models. In Advances in Economics and Econometrics. 10th World Congress, Vol. 3, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, 245 –95. Cambridge University Press.

Belloni, Alexandre, Chernozhukov, Victor and Hansen, Christian B. 2014a. High-dimensional methods and inference on structural and treatment effects. The Journal of Economic Perspectives 28(2), 29-50.

Belloni, Alexandre, Chernozhukov, Victor and Hansen, Christian, B. 2014b. “Inference on Treatment Effects after Selection among High-Dimensional Controls,” Review of Economic Studies 81 (2): 608-650.

Benabou, Roland, 2000. Unequal societies: Income distribution and the social contract. American Economic Review 90, 96-129.

Boix, C. (2003). Democracy and redistribution. New York: Cambridge University Press.

Chen, Shaohua, and Martin Ravallion, 2010, The developing world is poorer than we thought, but no less successful in the fight against poverty." The Quarterly Journal of Economics 125, 1577-1625.

Chernozhukov, V., Hansen, C. and Spindler, M., 2015. Post-selection and post-regularization inference in linear models with many controls and instruments. The American Economic Review, 105(5), pp.486-490.

Cowell, F., 2011. Measuring inequality. Oxford University Press.

Deininger, Klaus, Squire, Lyn, 1998. New ways of looking at old issues: inequality and growth. Journal of Development Economics 57, 259–287.

Delbianco, F., Dabús, C. and Caraballo, M.Á., 2014. Income inequality and economic growth: New evidence from Latin America. Cuadernos de Economía, 33(63), pp.381-398.

Dutt, Pushan and Mitra, Devashish. 2008. “Inequality and the Instability of Polity and Policy,” Economic Journal, 118(531): 843-860.

Easterly, William, 2001. The middle class consensus and economic development. Journal of Economic Growth 6, 317–336.

Easterly, William 2007. Inequality does cause underdevelopment: Insights from a new instrument, Journal of Development Economics 84, 755-776.

Engermann, Stanley, Sokoloff, Kenneth, 1997. Factor endowments, institutions, and differential paths of growth among new world economies: a view from economic historians of the United States. In: Haber, Stephen (Ed.), How Latin America Fell Behind. Stanford University Press, Stanford CA.

Forbes, Kristin, 2000. A reassessment of the relationship between inequality and growth. American Economic Review 90, 869–887.

Galor, Oded, 2011. Inequality, human capital formation and the process of development, In: Eric A. Hanushek, Stephen Machin and Ludger Woessmann (Ed.) Handbook of the Economics of Education, North Holland, 2011.

Galor Oded, Moav Omer, Vollrath, Dietrich, 2009. Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. Review of Economic Studies 76, 143-179.

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

Ghatak, Maitresh and Jiang, N. H. (2002). A simple model of inequality, occupational choice, and development, Journal of Development Economics, 69, 205–26.

Gimpelson V, Treisman D. 2017. Misperceiving inequality. Economics & Politics. 00:1–28.

Glaeser, Edward, Scheinkman, Jose, and Shleifer, Andrei 2003. The Injustice of Inequality, Journal of Monetary Economics, 50, 199-222.

Glaeser, E.L., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A., 2004. Do institutions cause growth? Journal of Economic Growth, 9(3), pp.271-303.

Haughton, Jonathan Henry, and Shahidur R. Khandker. 2009. Handbook on poverty and inequality. World Bank Publications, 2009.

Hebiri, M. and Lederer, J., 2013. How correlations influence lasso prediction. IEEE Transactions on Information Theory, 59(3), pp.1846-1854.

Jellema, Jon, and Gérard Roland. 2011. Institutional clusters and economic performance. Journal of Economic Behavior & Organization 79.1-2: 108-132.

Kaufmann, Daniel, Kraay, Aart, Mastruzzi, Massimo, 2009. Governance matters VIII: aggregate and individual governance indicators, 1996-2008. World Bank Policy Research Working Paper 4978.

Leamer, E.E., 1985. Sensitivity analyses would help. The American Economic Review, 75(3), 308-313.

Lee, D.J. and Son, J.C., 2016. Economic growth and income inequality: evidence from dynamic panel investigation. Global Economic Review, 45(4), pp.331-358.

Levine, Ross E. and Renelt, David, 1992. “A Sensitivity Analysis of Cross-Country Growth Regressions.” American Economic Review, 1992, 82(4), pp. 942–63.

Lupu, N. and Pontusson, J., 2011. The structure of inequality and the politics of redistribution. American Political Science Review, 105(2), pp.316-336.

Menezes, Carmen, Geiss, Charles, Tressler, John, 1980. Increasing downside risk. American Economic Review 70, 921-932.

Milanovic, Branko. 2002. True world income distribution, 1988 and 1993: First calculation based on household surveys alone, The Economic Journal 112, 51-92.

Moav, Omer. 2005. Cheap Children and the Persistence of Poverty. The Economic Journal, 115: 88–110.

Mookherjee, Dilip and Ray, Debraj 2003. ‘Persistent inequality’, Review of Economic Studies, vol. 70, 369–93.

Perotti, Roberto, 1996. Growth, income distribution, and democracy: what the data say. Journal of Economic Growth 1,149–188.

Persson, Torsten, Tabellini, Guido, 1994. Is inequality harmful for growth? American Economic Review 84, 600–621.

Piketty, Thomas 2014. Capital in the 21st Century. Cambridge: Harvard University.

Ravallion, M., 2012. Why don't we see poverty convergence?. The American Economic Review, *102*(1), pp.504-523.

Ravallion, Martin, 2003. The debate on globalization, poverty and inequality: Why measurement matters, International Affairs 79, 739-753.

Ravallion, M. and Chen, S., 2011. Weakly relative poverty. Review of Economics and Statistics, 93(4), pp.1251-1261.

Sachs, Jeffrey, Warner, Andrew, 1997. Fundamental sources of long-run growth. American Economic Review Papers and Proceedings 87, 184–188.

Sala-i-Martin, Xavier. 1997. “I Just Ran 2 Million Regressions.” American Economic Review, 87(2), pp. 178–83.

Sala-i-Martin, Xavier, Doppelhofer, Gernot, Miller, Ronald I. 2004. “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” American Economic Review, 94(4): 813-835

Sarsons, Heather 2015. Rainfall and conflict: A cautionary tale. Journal of Development Economics 115, 62–72.

Stock, James H., Yogo, Motohiro, 2004. Testing for weak instruments in linear IV Regression. Technical Working Paper 284.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

Vapnik, V., 2013. The nature of statistical learning theory. Springer Science & Business Media.

Voitchovsky S. 2005. Does the profile of income inequality matter for economic growth? Distinguishing between the effects of inequality in different parts of the income distribution. Journal of Economic Growth 10(3): 273–296

World Institute for Development Economics Research (WIDER), 2000, World Income Inequality Database, Version 1.0, User Guide and Data Sources.

Yitzhaki, S. 1983. On an extension of the Gini inequality index. International Economic Review, 24 (3):617-628.

Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), pp.301-320

Appendix 1

Table A1: Income Distribution Measures

Measure	Description
Headcount poverty at $2.00 a day (%)	Percentage of the population below $2 a day poverty line
Poverty gap index at $2.00 a day (%)	Extent to which individuals on average fall below the $2 poverty line, as percentage of poverty line
Poverty severity index at $2.00 a day (%)	Average of square of poverty gaps
Headcount poverty at $1.25 a day (%)	Percentage of the population below $1.25 a day poverty line
Poverty gap index at $2.00 a day (%)	Extent to which individuals on average fall below the $1.25 poverty line, as percentage of poverty line
Poverty severity index at $2.00 a day (%)	Average of square of poverty gaps
Sen index	The Sen index is a function of the headcount ratio, the mean income of the poor and the Gini coefficient of the poor.
Relative poverty measure (40% median)	Percentage of population with incomes below 40% of median income
Relative poverty measure (50% median)	Percentage of population with incomes below 50% of median income
Relative poverty measure (60% median)	Percentage of population with incomes below 60% of median income
Relative poverty measure (40% mean)	Percentage of population with incomes below 40% of mean income
Relative poverty measure (50% mean)	Percentage of population with incomes below 50% of mean income
Relative poverty measure (60% mean)	Percentage of population with incomes below 60% of mean income
Atkinson-Bourguignon poverty measure	Percentage of population with incomes below a hybrid poverty line combining World Banks' $2 a day line and relative poverty line of 37% below mean income
Atkinson-Bourguignon poverty measure	Percentage of population with incomes below a hybrid poverty line combining World Banks' $1.25 a day line and relative poverty line of 37% below mean income
Gini index	Twice the difference between the ordinate of the line of perfect equality and the ordinate of the Lorenz curve
Mehran index	Difference between the ordinate of the line of perfect equality and the ordinate of the Lorenz curve weighted by 1- p_i, See Yitzhaki (1983)
Piesch index	Difference between the ordinate of the line of perfect equality and the ordinate of the Lorenz curve weighted by p_i, See Yitzhaki (1983)
Atkinson, $\varepsilon = 0.5$	Atkinson measure with $\varepsilon = 0.5$

Atkinson, ε = 1	Atkinson measure with ε = 1
Atkinson, ε = 2	Atkinson measure with ε = 2
Coefficient of variation	Coefficient of variation of income
Relative mean deviation	Absolute mean deviation divided by the absolute value of the mean
Logarithmic standard deviation	Standard deviation of log income
Mean log deviation, GE(0)	Generalized entropy measure with α = 0; also mean log deviation
Theil index, GE (1)	Generalized entropy measure with α = 1
Half of coefficient of variation, GE(2)	Generalized entropy measure with α = 0; also half of the coefficient of variation
D1	Income share of decile 1 (in percentage)
D2	Income share of decile 2 (in percentage)
D3	Income share of decile 3 (in percentage)
D4	Income share of decile 4 (in percentage)
D5	Income share of decile 5 (in percentage)
D6	Income share of decile 6 (in percentage)
D7	Income share of decile 7 (in percentage)
D8	Income share of decile 8 (in percentage)
D9	Income share of decile 9 (in percentage)
D10	Income share of decile 10 (in percentage)

Table A2: Sample of Countries

Algeria	Ghana	Pakistan
Argentina	Greece	Panama
Armenia	Guatemala	Paraguay
Australia	Honduras	Peru
Austria	Hong Kong SAR, China	Philippines
Azerbaijan	Hungary	Poland
Bangladesh	India	Portugal
Belarus	Iran, Islamic Rep.	Romania
Belgium	Ireland	Russian Federation
Bolivia	Israel	Rwanda
Bosnia and Herzegovina	Italy	Senegal

Brazil	Jamaica	Singapore
Bulgaria	Japan	Slovak Republic
Canada	Jordan	Slovenia
Chile	Kazakhstan	Spain
China	Korea, Rep.	Sri Lanka
Colombia	Kyrgyz Republic	Sweden
Costa Rica	Latvia	Switzerland
Côte d'Ivoire	Lesotho	Tajikistan
Croatia	Lithuania	Thailand
Cyprus	Luxembourg	Trinidad and Tobago
Czech Republic	Macedonia, FYR	Tunisia
Denmark	Madagascar	Turkmenistan
Dominican Republic	Malaysia	Uganda
Ecuador	Mexico	Ukraine
El Salvador	Moldova	United Kingdom
Estonia	Morocco	United States
Finland	Netherlands	Uruguay
France	New Zealand	Uzbekistan
Georgia	Nigeria	Venezuela
Germany	Norway	Zambia

Table A3: Control Variables from Sala-i-Martin et al (2004)

Variable	*Description*
Absolute Latitude	Absolute latitude
African Dummy	Dummy for Sub-Saharan African countries
Air Distance to Big Cities	Logarithm of minimal distance (in km) from New York, Rotterdam or Tokyo
Average Inflation 1960-90	Average inflation rate between 1960 and 1990
British Colony Dummy	Dummy for former British colony after 1776
British legal origin	Dummy for common law countries
Capitalism	Degree Capitalism index
Civil Liberties	Index of civil liberties index in 1972

Colony Dummy	Dummy for former colony
Defence Spending Share	Average share public expenditures on defence as fraction of GDP, 1960-65
East Asian Dummy	Dummy for East Asian countries.
English Speaking Population	Fraction of population speaking English
Ethnolinguistic Fractionalization	Average of five different indices of ethnolinguistic fractionalization
European Dummy	Dummy for European economies
Fertility in 1960s	Fertility in 1960s
Fraction Buddhist	Fraction of population Buddhist in 1960
Fraction Catholic	Fraction of population Catholic in 1960
Fraction Confucian	Fraction of population Confucian in 1960
Fraction GDP in Mining	Fraction of GDP in mining
Fraction Hindus	Fraction of population Hindu in 1960
Fraction Muslim	Fraction of population Muslim in 1960
Fraction of Land Area Near Navigable Water	Proportion of country's land area within 100 km of ocean or ocean-navigable river
Fraction of Tropical Area	Proportion of country's land area within geographical tropics
Fraction Orthodox	Fraction of population Orthodox in 1960
Fraction Population In Tropics	Proportion of country's population living in geographical tropics
Fraction Population Less than 15	Fraction of population younger than 15 years in 1960
Fraction Population Over 65	Fraction of population older than 65 years in 1960
Fraction Protestants	Fraction of population Protestant in 1960
Fraction Speaking Foreign Language	Fraction of population speaking foreign language
Fraction Spent in War 1960-90	Fraction of time spent in war between 1960 and 1990
French legal origin	Dummy for civil law countries
Gov. Consumption Share 1960s	Share of expenditures on government consumption to GDP in 1961
Government Share of GDP in 1960s	Average share government spending to GDP between 1960–1964
Higher Education 1960	Enrolment rates in higher education
Hydrocarbon Deposits in 1993	Log of hydrocarbon deposits in 1993
Interior Density	Interior (more than 100 km from coastline) population per interior area in 1965
Investment Price	Average investment price level between 1960 and 1964 on PPP basis
Land Area	Area in square kilometres
Landlocked Country Dummy	Dummy for landlocked countries
Latin American Dummy	Dummy for Latin American countries
Life Expectancy in 1960	Life expectancy in 1960
Malaria Prevalence in 1960s	Index of malaria prevalence in 1966

Nominal Government GDP Share 1960s	Average share of nominal government spending to nominal GDP, 1960-64
Oil Producing Country Dummy	Dummy for oil-producing country
Openness measure 1965-74	Ratio of exports plus imports to GDP, averaged over 1965 to 1974
Outward Orientation	Measure of outward orientation. Levine and Renelt (1992)
Political Rights	Political rights index
Population Density 1960	Population per area in 1960
Population Density Coastal in 1960s	Coastal (within 100 km of coastline) population per coastal area in 1960
Population Growth Rate 1960-90	Average growth rate of population between 1960 and 1990
Population in 1960	Population in 1960
Primary Exports 1970	Fraction of primary exports in total exports in 1970
Primary Schooling in 1960	Enrolment rate in primary education in 1960
Public Education Spending Share in GDP	Average share public expenditures on education as fraction of GDP, 1960-65
Public Investment Share	Average share of expenditures on public investment as fraction of GDP, 1960-65
Real Exchange Rate Distortions	Real exchange rate distortions
Religion Measure	Religion measure
Revolutions and Coups	Number of revolutions and military coups
Size of Economy	Logarithm of aggregate GDP in 1960
Socialist Dummy	Dummy for countries under Socialist rule for considerable time during 1960-95
Spanish Colony	Dummy variable for former Spanish colonies
Square of Inflation 1960-90	Square of average inflation rate between 1960 and 1990
Terms of Trade Growth in 1960s	Growth of terms of trade in the 1960's
Terms of Trade Ranking	Terms of trade ranking
Timing of Independence	Timing of national independence measure: 0 if before 1914; 1 if between 1914 and 1945; 2 if between 1946 and 1989; and 3 if after 1989
Tropical Climate Zone	Fraction tropical climate zone
War Participation 1960-90	Indicator for countries that participated in external war between 1960-1990
Years Open 1950-94	Number of years economy has been open between 1950 and 1994. Sachs and Warner index

Appendix 2

In our analyses so far, we use the Worldwide Governance Indicators from Kaufmann, Kraay, and Mastruzzi (2009). However, as Glaeser, La Porta, Lopez-de-Silanes, and Shleifer (2004) and Jellema and Roland (2011) argue, many institutional indicators, including the ones we use, are outcome measures and confound constraints on political and executive decision making (e.g., a democracy with checks and balances) with choices made by political leaders (e.g., a dictator who faces few constraints but chooses good regulatory quality, intolerance for corruption, etc.) Additionally, such institutional measures may be transitory, reflecting recent political outcomes or subjective assessments that do not capture permanent and deep constraints on executive decision making. Therefore, as a robustness check, we go beyond the governance indicators and replace them with a host of institutional indicators. Based on Jellema and Roland (2011) we consider the following additional institutional indicators: Constraints on the Executive, Autocracy Score and Democracy Score from the Polity IV database; Civil Rights and Political Rights from Freedom House where high numbers indicate fewer civil and political rights; a corruption index from Transparency International with higher numbers indicating higher levels of corruption; a dummy variable that takes the value 1 for Proportional Electoral Rule from the Database of Political Institutions; a categorical variable that takes the value 2 for Parliamentary regimes, 1 for Assembly-elected President and 0 for Presidential systems; Indices of Legislative and Executive Competitiveness from the Database of Political Institutions. To facilitate comparison with the governance measure, we used data from 2002.

We used the double-selection procedure to estimate equation (4) and replicated Column (5) in Table 5B for these ten additional institutional measures. These results are presented in Table A5 below. With these alternate measures of institutions, we again find that poverty only has a weak effect on institutional quality. Poverty does not matter for constraints on the executive, for democracy and autocracy scores, political rights and structural features of the political system. Poverty reduces civil liberties and is associated with higher levels of corruption.

Table A4: Additional Development Outcomes

	Life Expectancy	*Infant Mortality*	*Maternal Mortality*	*Happiness Index*	*Trust*	*Primary education*
Gini	0.004	-0.407**	-5.806***	-0.010	-0.669**	0.270
	(0.065)	(0.189)	(1.853)	(0.018)	(0.278)	(0.211)
Poverty	-0.061**	0.260***	2.121**	-0.007	0.080	-0.402***
	(0.026)	(0.083)	(1.020)	(0.008)	(0.103)	(0.096)
R^2	0.96	0.95	0.89	0.78	0.74	0.87
N	63	60	61	46	46	51

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%;
All columns include a constant and controls chosen by the double-selection procedure (not shown)

Table A5: Alternate Measures of Institutional Quality

	Constraints on executive	*Autocracy Score*	*Democracy Score*	*Civil Liberty*	*Political Rights*	*Corruption*	*Parliamentary System*	*Proportional Representation*	*Legislative Competitive ness*	*Executive Competitive ness*
Gini	0.081***	-0.077**	0.160***	-0.016	-0.001	0.023	-0.002	-0.002	0.073*	0.080**
	(0.029)	(0.032)	(0.051)	(0.021)	(0.018)	(0.030)	(0.008)	(0.008)	(0.039)	(0.038)
Poverty	-0.005	0.004	0.000	0.014*	0.005	0.029**	0.003	0.001	-0.010	-0.015
	(0.011)	(0.012)	(0.020)	(0.008)	(0.007)	(0.014)	(0.003)	(0.003)	(0.014)	(0.011)
R^2	0.63	0.64	0.72	0.84	0.85	0.83	0.72	0.56	0.28	0.32
N	61	61	61	60	62	54	62	60	62	62

Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%;
All columns include a constant and controls chosen by the double-selection procedure (not shown)