The Determinants of Income Inequality: a Bayesian Approach to Model Uncertainty

Mark Hopkins
Department of Economics
Gettysburg College
300 N. Washington St., Box 391
Gettysburg, PA 17325

October 11, 2004

Abstract

This paper explores the determinants of income inequality using a cross-country panel of Gini coefficients. Economic theory suggests a wide range of potential mechanisms generating inequality, but there is little consensus regarding the most relevant, nor is there a generally accepted empirical specification for use in Gini coefficient regressions. Drawing appropriate inference from the cross-country data requires formal recognition of this uncertainty in the search for model specification. This paper adopts a Bayesian approach to inference that that averages across models using Bayesian posterior model probabilities to assess the marginal impact of specific variables across a range of potentially likely "true" model specifications. The resulting posterior mean and standard error of each coefficient reflects more accurately our uncertainty regarding both parameter values and model specification.

1 Introduction

Over the last decade, a large empirical literature emerged investigating the determinants of the distribution of per-capita incomes across countries. The determinants of income distributions *within* countries have, by comparison, received less attention. Although much has been written on the subject of income inequality, empirical work in the area has traditionally focused on accounting for trends in household level data within specific countries, rather than addressing deeper questions of the long-run macroeconomic determinants from a comparative international perspective.

Figure 1 highlights two interesting features of the international pattern of inequality (the source of the data is detailed in Section 3). First, there is significant variation in levels of income inequality across countries, ranging from particularly high levels in Africa and Latin

America to fairly low levels in Eastern and parts of Western Europe. Second, within countries, inequality displays remarkable persistence relative to the cross-sectional variation.[1] Even with notable increases or decreases in inequality in some countries (including a rapid rise in the past decade among countries transitioning from socialism to market-based economies), high (or low) inequality regions have sustained high (or low) levels of inequality over most of the post-war period. In general, the range of inequality in the time series of a single country averages an order of magnitude less than that of a cross-section at a given point in time.[2] These empirical phenomena merit a search for an explanation.

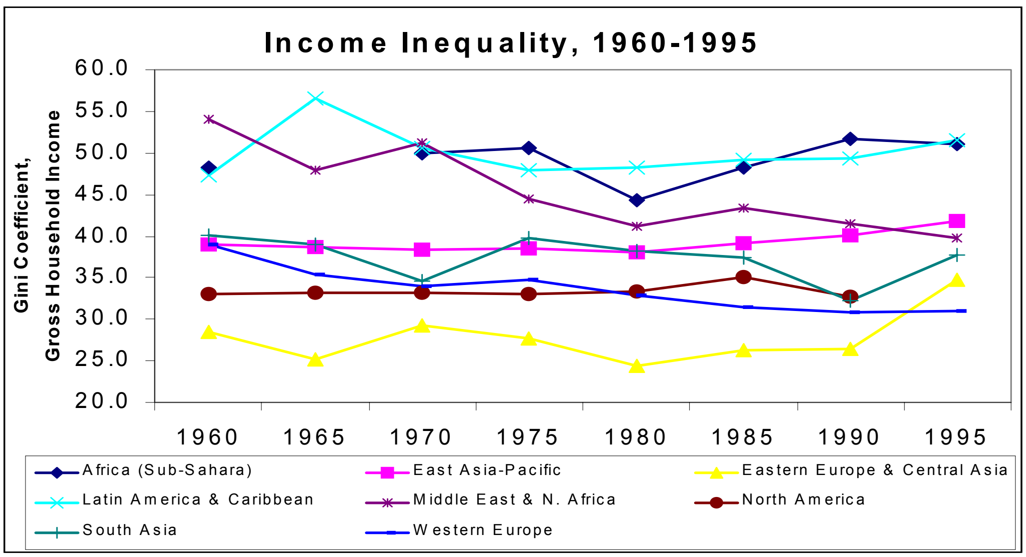


Figure 1: Data represent regional averages of national Gini coefficients measuring gross household income, or adjusted to approximate gross household income as described in Section 3 of the text. Inequality data are available only as an unbalanced panel, so changes in sample composition are also partially responsible for period to period movements (particularly in less developed regions).

The imbalance in research between the determinants of growth and the determinants of inequality most likely reflects two factors. First, growth in per-capita income has traditionally been seen as the central issue in development work, with distributional issues at best a

[1] For clarity, Figure 1 displays regional averages rather than country time series. It should be noted that, although these averages may mask some time-series fluctuations in individual countries, they also generate some spurious movements because of compositional changes resulting from the unbalanced country panel data.

[2] The support of Gini coefficients across countries is roughly from 20 to 60. By comparison, the reported rise in the U.S. Gini coefficient from 1980 to 1990 was from 35 to 38 (the lowest value since 1947 was 33.5, in 1968). In the sample of Gini coefficients used in this paper, the standard deviation across countries is roughly 10, versus about 3.3 within-countries. Li, Squire, and Zou (1998) present similar findings based on the **DS** data alone.

secondary concern, and at worst outside the bounds of the "Pareto criterion" (and by extension outside the bounds of an acceptable reserach agenda for economists: see, e.g., Feldstein (1999)). Although attributing normative significance to differences in inequality certainly requires caution, seeking to understand the positive economics governing the distribution of national income is a pursuit that dates back to classical economists of the 18th and 19th centuries.

The second reason inequality research has lagged growth research has been the historical scarcity of reliable distributional information across more than a few countries. An important contribution in this regard was made by Deininger and Squire (1996) [henceforth **DS**] of the World Bank, who compiled summary inequality statistics based on income surveys conducted across a range of countries over the post-war period. Using the **DS** data set, a number of papers addressing cross-country issues of inequality soon emerged, including Bénabou (1996), Forbes (2000), Barro (1999), Li, Squire, and Zou (1998)and Bannerjee and Duflo (1999). The focus of most of these papers has been on the role of inequality as a determinant of growth, however, not the determinants of inequality itself. Where this question has been explored, the approaches taken have varied substantially, reflecting the large number of questions being asked and the lack of consensus over an appropriate statistical model for the Gini coefficient.

1.1 Empirical Evidence on Inequality

The appropriate empirical methodology for understanding income inequality data across countries has received little attention. Within countries, where information on the joint distribution of household incomes (y) and associated characteristics (x) is available, a standard variance decomposition

$$Var(y) = Var_x(E(y|x)) + E_x(Var(y|x))$$

can be used to address the question of "how much inequality can be explained" by variance in group characteristics, relative to (unexplained) "within-group" variance (e.g., Cowell and Jenkins (1995)). The macroeconomic literature has taken a very different approach to "explaining" inequality, however, reflecting the aggregate nature of the data. Empirical work studying inequality across countries has generally taken an *ad hoc* approach, estimating simple linear reduced form specifications

$$Inequality_j = \sum_{k=1}^{p} \boldsymbol{\beta}_k X_{kj} + \varepsilon_j \tag{1}$$

by regressing p characteristics of economies j on a summary inequality statistic such as the Gini coefficient.

In this context, explanations of inequality are focused not on aspects of individual heterogeneity within countries, but on differences across countries in macro aggregates. Relationships in the micro studies between income and characteristics such as age, education, and race are expected to appear as correlations between the Gini coefficient and demographic characteristics, the distribution of education over the population, and ethnic fractionalization. As discussed later, however, the assumption of both linear relationships and simple additive structures in this context is questionable.

Articles by Li, Squire, and Zou (1998) (henceforth **LSZ**), Gustaffson and Johansson (1999) (**GJ**), Bénabou (1996), and Barro (1999) are recent examples of work employing specifications like equation (1). The regressions in the first two papers represent specific "tests" of the predictions of various economic theories, while the latter two take a more investigative approach. Although all of the authors make use of the longitudinal data available, the focus of **LSZ** and Barro is variation in *levels* of inequality across countries, while that of **GJ** and Bénabou is explaining *changes* in inequality over time. A comparison of their basic regression results are presented in Table 1 (all Tables can be found at the end of the paper).

LSZ assess the empirical support for two specific hypotheses that have been proposed as explanations for inequality: poor credit markets and political economy.[3] They find evidence for both effects, using the ratio of M2 to GDP and a Gini coefficient for land use (proxying asset inequality) to test the effect of credit market imperfections, and an index of civil liberties and the initial (1960) level of secondary schooling to test for political economy effects. **GJ** test a wider range of theories, including the role of Kuznets-style structural adjustment (changes in employment shares in the industrial sector), import penetration, unemployment, inflation, and various political, institutional, and demographic variables. The authors' explicit specification search to look for "smoking guns" comes closest to the approach of this paper, although GJ search for a single "best" specification using classical hypothesis testing while this paper attempts instead to assess the relative probabilities of a range of model specifications. Using a fixed effects model, **BJ** find that increases in the industrial share of employment is associated with decreases in inequality within countries, and suggest the same may be true for trade unions and the public sector share of GDP as well. Their evidence also supports a positive association between the import share of GDP and inequality: given their sample is restricted to OECD countries, this supports the conventional general equilibrium logic of trade theory.

The papers by Bénabou and Barro focus primarily on the impact of inequality on growth, but each author devotes some attention as well to the process generating the Gini coefficient data. Bénabou's interest is in dynamics, looking for evidence of convergence in Gini coefficients. Running Galton-style regressions across countries, he finds that inequality is rising on average, but a significant negative correlation exists between initial levels of the Gini coefficient and the subsequent average rate of change. His results are not entirely robust to choice of time period, however, and test only for unconditional convergence: no attempt is made to control for the effect of other variables associated with inequality. Bénabou also makes minimal adjustment for heterogeneity in the observations due to survey techniques, which may

[3] As discussed in the theory section, there are several distinct "political economy" considerations to consider with regard to inequality. The description of the political economy mechanism described by LSZ is a story of institutional strength providing protections afforded to the poor from expropriation by the rich.

result in spurious findings of convergence to the mean due to measurement error.

Barro examines the cross-sectional pattern of inequality at four decade intervals between 1960 and 1990 using SUR estimation. Unlike LSZ or GJ, Barro's approach is directed more toward finding predictive associations among the variables than testing specific theories of inequality determination. Reporting a set of results from potential model specifications, Barro establishes reduced form correlations of the Gini coefficient with levels of education, dummies for Africa and Latin America, and per-capita income (finding the standard Kuznets curve), but little evidence of linear correlations with democracy or the rule of law.

A comparison of the authors' results is presented in Table 1. It is remarkable that four separate attempts to "explain" inequality present four entirely non-nested model specifications. The lack of any single benchmark framework for estimation complicates both our inference about coefficient estimates (which is conditional on the controls employed and the estimation method) and the consideration of additional covariates. The R^2 reported in Table 1 for each of the models run by LSZ, BJ, and Barro is roughly the same, with the covariates explaining about two-thirds of the variation in the Gini coefficient.[4] In terms of predictive ability, no model stands out as superior to the others.[5] What is needed, therefore, is a systematic way of incorporating the information derived from each of these regressions, in a way that respects our understanding that it is not unique.

1.2 Macroeconomic Theory and Model Choice

Two challenges faced by statistical work seeking to explain the cross-country pattern of income inequality are the difficulty posed by the breadth and complexity of the potential explanatory factors and the lack of a clear statistical model to use as the basis of for regression with aggregate data. Although economic theory suggests a wealth of potential macroeconomic factors by which aggregate inequality statistics can be influenced (these are reviewed in Section 2), tests of these theories requires identification of each of their effects through their reduced form relationship with summary inequality statistics, typically the Gini coefficient.[6] Dishearteningly, in some of the studies much of the cross-country variation in inequality is picked up by country- and/or region-specific dummy variables, suggesting that important sources of inequality are yet to be fully understood.

[4] The very low R^2 in Bénabou's specification can be attributed to (a) the lack of conditioning variables and (b) the fact that the dependent variable is no longer the level of the Gini, but the average *change* in the Gini between the first and last periods reported, which undoubtedly magnifies the noise-to-signal ratio given measurement error in the reported Gini coefficients and the persistence of inequality.

[5] These four studies differ in several respects other than the covariates employed, however, including the authors' choices regarding the data set used, the type of questions addressed, and the econometric techniques employed. These issues are discussed in Appendix 1 of the paper.

[6] For a fairly large subset of the countries with Gini coefficient data, quintile shares (the Lorenz curve data underlying the Gini coefficient) are also reported. However, the structural relationship between the Gini coefficient and quintile shares (which are themselves highly collinear) implies that, from a practical perspective, not much is gained in terms of additional information.

More troubling from the standpoint of policy is the fact that most model specifications adopted in the literature are non-nested, having few, if any, covariates in common. This complicates inference in three ways. First, when estimating coefficients on specific variables, the exclusion of variables found to be "significant" in other studies should raise concerns regarding bias in the reported estimates. Second, from the standpoint of prediction, a single "best approximation" to the unknown process determining inequality is preferable to a collection of estimates taken from smaller specialized models.[7] Third, when introducing new variables into the analysis, the lack of an appropriate benchmark raises the question "to which of the existing specifications should a new covariate be added?"

Much of the lengthy econometric literature on the subject of model selection is based on the principle of "nesting and testing," or estimating large, flexible models and then performing hypothesis tests on various restrictions to find the most constrained model not substantially at odds with the data.[8] Estimating highly flexible models is far from a panacea to addressing model uncertainty, however. In the practice of cross-country regressions, the "nest and test" approach is complicated by what **Mankiw (1995)** refers to as a "degrees of freedom problem": a wealth of covariates suggested by theory and a paucity of country observations. It also raises concerns about "overfitting," to arrive at specifications overly sensitive to the particular characteristics of the data sample. The ability to discriminate among competing theories of inequality is further hampered by the collinearity among many of the variables of theoretical interest. For these reasons, it is widely accepted that econometric model building should proceed from and be guided foremost by theoretical considerations.

Yet macroeconomic theory offers little guidance in specifying a basic statistical model for the Gini coefficient in the same way that, for example, the Solow model has guided regression specification in the growth literature (e.g., Mankiw, Romer, and Weil (1992) and Barro (1991)). Many theoretical models of inequality share little in common, with some focusing, for instance, on factor distributions and others factor prices. Moreover, many of the simplifications made to provide tractable theoretical frameworks compromise their applicability in guiding empirical specifications. In other cases the data required to test the theory appropriately (e.g., that regrading the distribution of wealth or individual characteristics) are unavailable in cross-country work. The predictions offered by economic theory for the set of observable international inequality data are therefore rather vague, suggesting little more than a set of candidate explanatory variables, a set much larger and more collinear than can reasonably tested with the available data.

[7] This point does not contradict arguments made later in favor of inference based on model averaging. The point is simply that we should consider estimates to be "better" when drawn from a more likely model. Informal combining of estimates is methodologically inappropriate, and raises the uncomfortable possibility of accounting for more than 100% of the variance in the Gini coefficient.

[8] Classical or "frequentist" hypothesis tests of model restrictions are based on comparisons of maximized likelihood functions and/or F-tests based on residual sum of squares. As shown in section 4, these are related in turn to Bayesian posterior model probabilities.

1.3 Bayesian Approaches to Inference under Uncertainty

The fundamental challenge in designing an empirical framework to study inequality is to admit appropriate inference from the data within the context of our uncertainty regarding appropriate model specification. However, under model uncertainty the econometrician's problem becomes two-fold: first a statistical model must be chosen, then estimation of parameters can be performed conditional on that model. As Western (1996) notes, however, this becomes "a highly inductive enterprise that engages the data in double duty....This type of uncertainty, which necessitates a data-driven model search is not accomodated by conventional statistics. When the same data are used for model selection and parameter estimation, the usual p-values and standard errors can be biased in the direction of positive findings."[9]

Many of the concerns regarding specification searches in cross-country regressions discussed so far have been raised previously in the context of the empirical growth literature (e.g. , Mankiw 1995, Levine and Renalt, 1992; Sala-i-Martin, 1997). Levine and Renalt search for those variables among a candidate set appearing in growth regressions that are robust (in the sense of Leamer (xxxx) to specification. Sala-i-Martin adopts a more forgiving standard, assessing degrees of "certainty" about each potential variable. Neither of these papers provide much in the way of clear guidance towards either a specific model, or appropriate parameter inference. Recent efforts to address such issues in the growth literature have centered around Bayesian Model Averaging (Fernández, Ley and Steel, 2001; Brock and Durlauf, 2001); following suggestions made by Leamer (1978). In the sociology literature, Western (1996) has proposed Bayesian model averaging techniques explicitly as a way of dealing with problems of vague theory. This paper builds on these ideas, setting out a Bayesian approach to model building to assess how macroeconomic and country characteristics affect the distribution of income.[10]

The discussion proceeds as follows. A brief overview of some of the theoretical determi-

[9] The heterogeneity of the country observations provides yet another dimension of uncertainty, as estimated relationships may be sensitive to the particular country sample used. This can complicate specification searches further since, with missing data, the set of observations available for estimation may change with the composition of the covariate vector. (For instance, Barro (1999) discusses the role of "rich country sample bias" in a discussion of the results of Forbes (2000).) In this respect, some level of robustness not only to model specification but also to sample selection is crucial to achieving appropriate statistical inference. The "correct" statistical model can be thought of as that which leads to exchangeability of the error terms, however, and this respect incorporation of parameter heterogeneity can be thought of as a subset of the larger question of model choice.

[10] In point of fact, there is nothing uniquely Bayesian to many of the ideas employed in this paper. Choice of model priors is rather primitive, with all permutations of the variables under consideration treated as being "equally probable." Many of the technical arguments and computational details can also incorporated into a frequentist framework. Burnham and Anderson (1998) derive likelihood-based model selection and model averaging tools using AIC-based weights, using arguments similar to those used by Raferty (1995) to motivate his BIC approximation to the posterior density. What is principally "Bayesian" in this paper is the exposition of methodology in terms of a posterior subjective probability distribution over models conditional on the data, which are more intuitive than arguments based on information theory regarding the expected Kullback-Leibler distance from the "true" model.

nants of income inequality is presented in Section 2 to motivate the set of explanatory variables adopted later in the paper. Construction of the data set employed in this paper is detailed in Section 3, discussing in particular the choices made in trading off quantity, quality, and consistency of the observations relative to the approach of other papers. Section 4 outlines the trade-offs involved in the model selection decision, and compares methods suggested in the both the classical and Bayesian literatures on specification search, noting the implications for inference. Section 5 presents a meta-analysis of the current literature, using model selection techniques to discriminate between the specifications of Li, Squire, and Zou (1998), Gustaffson and Johansson (1999), Bénabou (1996), and Barro (1999), then expanding the technique to evaluate a larger set of candidate variables. In Section 6 a Bayesian technique introduced by George and McCulloch (1993) and Smith and Kohn (1996) for the simultaneous selection of variables and functional specification is employed to explore the determinants of within-country movements in inequality. The results are contrasted with other techniques, including (forward and backward) stepwise regression procedures based on maximizing the Akaike information criterion, and model averaging is used to summarize the conditional distribution of the model parameters in a manner independent of specific model choice. Section 7 summarizes and concludes.

2 Economic Theories of Inequality

Research on international comparisons of inequality has been shaped in large part by the seminal address of Kuznets (1955), who hypothesized that, in the process of economic development from rural agrarianism to an urban industrial society, inequality would first increase and then decrease. This hypothesis preoccupied research for the next half century as researchers debated whether or not a "Kuznets curve" could be found in the data. Even today, the Kuznets hypothesis continues to drive empirical work on the determinants of inequality (e.g., Barro, 1999; Gustaffson and Johansson, 1999; and Li et. al., 1998)., , and). Nevertheless, theoretical work on inequality over the past decade has delivered a much richer menu of possible questions to be addressed by the empirical researcher.

The Kuznets hypothesis was originally motivated through a process of structural and demographic transition, but for most inequality measures this "inverse U-shape" pattern can arise when limited resources require the population be distributed across unevenly productive activities. Examples include matching between workers and new capital vintages (Jovanovic (1998)) or technologies, and between investors and financial intermediaries (Greenwood and Jovanovic (1990)). More generally, a Kuznets-type concave non-linearity can emerge, for a given set of returns to various activities, when the distribution of individuals over those activities changes.[11] Consider for a moment the Gini coefficient, the most common measure

[11] It is not necessary – or likely – that returns to various activities are fixed. All that is required is that the returns not vary too greatly relative to the distribution. In terms of a simple production function with two factors (e.g. unskilled and skilled labor), an elasticity of substitution greater than one will guarantee concavity.

of income inequality, which can be defined as

$$G(y) = \int \left(F(y) - \frac{1}{2} \right) y dF(y)$$

where $F(y)$ represents the distribution of incomes, y. If we assume that (otherwise homogeneous) agents are distributed over two activities (be it sectors, technologies, or capital vintages) L (low) and H (high), offering fixed incomes $y_L < y_H$, respectively, and we denote by λ the fraction assigned to activity H, then the Gini coefficient for the economy equals

$$G(y) = \left(\frac{y_H - y_L}{\mu} \right) \lambda \left(1 - \lambda\right) \tag{2}$$

where $\mu \equiv y_L + \lambda(y_H - y_L)$ is per-capita income.

For given incomes y_L and y_H, the Gini coefficient is a concave function in λ, while per-capita income is a linear function of λ. Changes in λ will therefore trace out a "Kuznets curve" (an inverse-U shaped relationship) between inequality G and average incomes μ. A conventional interpretation of λ is the share of workers in the "modern sector." Alternatively we can think of λ is the "skilled" share of the population. While equation (2) clearly presents a very simplified formula for the Gini coefficient, it has important implications for thinking about the specification of inequality regressions. In particular, many variables used in inequality regressions, such as "average years of schooling," are presumably used because they have been common in growth regressions. However, equation (2) suggests that linear regressions of average years of schooling (which is linearly related to λ) on the Gini G will be mis-specified without a quadratic term, since the true relationship is non-linear in λ.

2.1 Economic Growth and Development

Although the Kuznets curve was once considered to be an empirical regularity, additional evidence has brought this into question, particularly with regard to time series behavior of inequality within countries (which, as noted in the introduction, shows remarkable persistence).[12] Of course, in evaluating the role of "development" on inequality, is important to note that Gini coefficients are bounded and scale invariant by construction and appear stationary in a time series, while neither is true of the level of income. The ubiquitous practice of regressing the (stationary) Gini coefficient on the (non-stationary) level of income is therefore somewhat questionable. Two approaches to this issue are taken. The first is to include with per-capita income more direct measures of λ, the factors presumed to be driving the Kuznets dynamics. Examples include the share of the skilled population; the share in agriculture, industry, and services; and/or the share in urban areas. As highlighted equation (2), these

[12] Li, Squire, and Zou (1998) do not find evidence for a curve in the time-series, although Barro (1999) claims to have done so. Testing the hypothesis of a Kuznets curve in a time series ignores certain methodological problems, of course. Since the process of development has no clear "beginning" or "end" it is not immediately clear what statistical relationship between levels of inequality and per-capita income should result.

terms should each be allowed to enter with a non-linearity. Since including the level of income is only a concern from the standpoint of time-series variation, the second approach is to distinguish between average income levels across countries and changes within countries over time.

Changes in λ are not the only way that that economic development may affect inequality, however: relative rates of return are likely to change as well. Many economists attribute the rise in earnings inequality observed in the US to the same technological innovations credited for rapid U.S. growth (see, e.g., Acemoglu (2002)). Moreover, innovations are typically both local and proprietary – they bring new opportunities and rewards to entrepreneurs and patent holders before the public at large – so a temporary rise in inequality may accompany periods of rapid growth. These facts suggest that in the few decades over which we have data, short-run correlations between inequality and the *rate of growth* may be more relevant features of a time series than long-run Kuznets relationships between inequality and the level of income.

There are other potential channels through which development can affect inequality, of course, including the strengthening of governments and other social institutions. Although altruistic behavior and the ubiquity of beliefs in "social justice" suggest that humans value both their own and their neighbors' well-being, it is not unreasonable to assume that the income elasticity of the latter is greater than the former. As a result, development may be accompanied by expanded government provision of public services benefiting the poor and other progressive policies.[13]

2.2 Demographics

The "true" degree of inequality in lifetime earnings is most likely overstated by household income Gini data because of the empirical age-earnings profile. Differences in the slope of the age-earnings profile and the demographic distribution of ages each may explain some fraction of the variation in measured inequality. One method of controlling for the age profile of a country is by utilizing data on the share of the non-working age population under age 15 and over age 65.

In addition to the static age distribution, inequality may be correlated with population dynamics as well, as both fertility choices and mortality rates may affect economic subsequent inequality. If less educated parents may have a larger number of children – due to the low opportunity cost of their time, high infant mortality rates, or as a substitute for other retirement assets – and if the children of less educated parents are on average also less educated themselves, then higher fertility rates are likley to be associated with a larger supply of unskilled labor relative to skilled labor, supressing the growth in the relative wages of the unskilled and widening inequality. Mortality rates may also be expected to play a role in observed inequality, most notably through education decisions. Human capital investment

[13] One must of course consider the usual caveat that, due to reverse causation, the richest societies may be those with the least redistributive taxation.

involves up-front fixed costs and an annuity return in terms of higher wages during the years worked, which implies that a greater probability of early death, injury or illness lowers the average expected lifetime return to education. Education may be an important factor in reducing mortality, however. To the extent that the poor in a country are more likely to suffer from a vicious cycle of short life expectancy and less education than the rich, this may explain in part persistent differences in inequality across countries.

2.3 International Trade & Investment

The impact of globalization on the level of inequality in both developing and developed countries is one of the most contentious issues facing policy makers today. Conventional wisdom on the relationship between globalization and inequality is rooted in general equilibrium trade theory and the notion that international trade equalizes imbalances in relative factor scarcity across countries, putting upward pressure on returns to relatively abundant factors and downward pressure on relatively scarce factors. A standard interpretation is that freer trade will cause the returns to capitalists and educated workers to fall relative to unskilled workers in poor countries and to rise in rich countries.

Trade in final goods is only one component of globalization, however, and the role of intermediate goods trade and foreign direct investment in "outsourcing" decisions complicates any attempts at making any broad predictions. Feenstra and Hanson (1996) show how movements of capital (FDI) from the North to the South induces a rise in the skill premium in both regions because the marginal industries that move to the South are the least skill intensive in labor demand relative to Northern industries, but the most skill intensive relative to Southern industries. Zhu and Trefler (2001) derive similar results using technological catch-up rather than FDI to shift industries to the South.

There is no consensus on how best to quantify globalization: several "openness" measures have been suggested in the literature, and each has its strengths and weaknesses. Most common is a volume of trade measurement, the share of exports plus imports to GDP. This measure hardly seems to capture a country's policy stance, although it does seem a reasonable proxy for a country's integration and dependence on world factor markets. Including an additional term for the interaction with the income level allows a rough test of the simple 2x2x2 general equilibrium logic.

It should be noted that all of the theoretical mechanisms described above focus on factor price effects, and thus most appropriately describe a "medium-run" scenario: sufficient time has passed for general equilibrium effects to run their course, but not so much as to violate the assumption of fixed factor endowments. Terms-of-trade shocks are likely to be an important factor in understanding movements in inequality within this periodicity.

Figure 2 examines the evolution of inequality according to countries' principal type of exports. As in Figure 1, the unbalanced panel generates some noise in period to period movements due to sample composition changes, but the data do suggest a few clear trends.

First, there was a noticeable decline in income inequality in oil exporting nations between 1960 and 1980 (particularly during the 1970s) – a period of time in which oil prices were particularly high.[14] This suggests an interesting and perhaps powerful connection between income inequality and terms-of-trade shocks. There appears to be a clear trend toward rising inequality in other (non-oil) primary commodity exporters, however, while manufacture exporting countries enjoyed a secular decline in inequality (although this trend was suddenly reversed between 1990 and 1995). Diversified and services exporters do not appear to have significantly different levels of inequality than they did in 1960. In short, it appears that both long-run (cross-sectional) correlations between openness and inequality, and short-run (within countries) changes in terms-of-trade may be important determinants of inequality.

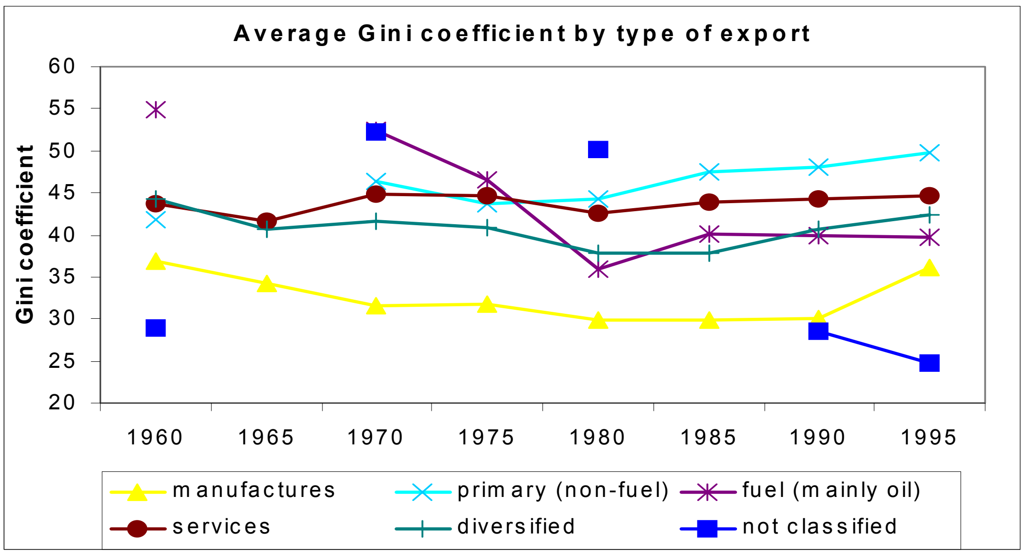


Figure 2: Gini coefficient means by principal exports. Missing observations represent periods with insufficient data. Sample sizes for periods 1960-1995 were, respectively, 39, 32, 57, 62, 77, 80, 102, 66.

2.4 Non-covexities, history, and the role of institutions

As noted in the introduction, there appear to be at least two clear stylized facts regarding the cross-country pattern of inequality:

1. Aggregate measures of inequality vary little in a time series relative to a cross-section, and there appears to be little systematic long-run movement in most countries relative

[14] This corresonds with the fall in inequality in the Middle East and North Africa shown in Figure 1, although the trend among oil-exporters as a group appears stronger in the 1970s than that of Middle Eastern countries as a group, suggesting that the relevant indicator may be trade rather than geography.

to short-run fluctuations.

2. Most of the variance across countries exists across "groups" of countries. This argument is true in particular both for groups defined by geographical region and/or by the origin of legal systems & institutions.

The first of these stylized facts is consistent with a class of models in which multiple long-run steady-states exist for the income distribution due to non-convexities in the investment technology (c.f., Galor and Zeira, 1993). As a result, "history matters" in these models: the current distribution of individual investment and income is linked directly by a process of intergenerational bequests to some "initial" wealth distribution. Inequality is more likely to exist – and to persist – when private credit markets are poorly developed and significant frictions exist.

A rigorous evaluation of this hypothesis requires international data on the distribution of asset wealth, which is not currently available. In a deeper sense, however, these models imply that the current distribution of income is simply a manifestation of some "original" asset distribution with the implication that, to explain the current cross-country pattern of inequality, we should look to the historical structure of government (feudal, aristocratic, or democratic) and the legal system, particularly with respect to the assignment and protection of individual property rights. Sokoloff and Engerman (2000) attribute present-day disparities in levels of inequality between North and South America to differences in the property rights regimes instituted under colonialism. This formulation of the distributional path-dependency argument is consistent with both stylized facts listed above, explaining both the cross-sectional variance and time-series persistence, and why inequality is correlated with regional groups and legal systems.

It should be noted, however, that while history provides many examples of institutions shaping the income distribution, the net effects are far from obvious. Theory can offer some *a priori* guidance on what the effect of particular institutional characteristics should be, however. Institutions that guarantee property rights are likely to foster investment and growth. This may also engender a widening in the income distribution, however, given the inherent riskiness of investment and the protection of the rewards of the winners from expropriation by the losers. On the other hand, institutions that guarantee civil liberties help prevent the exploitation of the poor by the wealthy elite in economic bargaining. Taken further, institutions that deliver political rights uniformly across the public can generate pressure for populist redistributive policies.

Although the case for including separate institutional variables for property rights, civil liberties, and political rights can be made in theory, complications arise in practice from collinearity among these variables in the data. Therefore, institutional measures used in this paper are restricted to an index of economic freedoms on the one hand – a measure of respect for property rights, taken from Levine, Loayza, and Beck (2000) – and political freedoms (alternatively, "democracy") on the other.[15]

Although the impact of these variables is most likely to be seen in the long-run, and therefore most evident in a cross-section, institutional shocks may also affect inequality in the short-run. A number of variables measuring institutional risk and/or instability have been suggested in the political science literature, and have been used by economists in economic growth research. These include political assassinations, revolutions and coups, and leadership and/or constitutional changes.

2.5 Final considerations

The goal of the taxonomy of inequality theories presented above is to generate an *a priori* defensible set of candidate variables that can be employed in cross-country regressions to uncover patterns in the data that might suggest "causes" of income inequality. No claim is made regarding either the completeness or the definitiveness of the taxonomy. Some in particular may find the lack attention to public finance objectionable. Since the unit of observation is the nation state, differences in tax and expenditure policies across governments seems a natural explanation for why inequality differs across countries involve tax and social policies. (Although our Gini coefficients are based on gross household income, they should in theory reflect distortions in saving and labor decisions due to taxes, as well as the level of public transfers and public goods expenditures). In many cases, in fact, the time series behavior of inequality within specific countries can be convincingly explained by the nature of political events and policies in that country.[16]

The reason for the lack of attention to redistributive and social policies is two-fold. The first is practical: international data on tax and transfers is limited, and thus limits the extent to which we can draw inference from a fairly representative sample of countries across regions and income levels. Second, the approach taken of subsuming public finance within broader issues of development and institutional quality reflects comparative advantage. The existing literature on the comparative study of inequality is dominated by public and labor economists whose analyses of the role of public policies in shaping inequality is much more comprehensive than is possible here, given the number of countries examined and the level of abstraction taken with respect to each. As always, the justification for such abstraction is to highlight broad trends that apply generally to many countries, rather than attributing all variation in inequality to specific events – to see the "forest" rather than the trees.[17]

---

[15] Barro (1999) uses an average of the civil liberties and political rights data from Freedom House (www.freedomhouse.com) to provide a measure of political freedoms he labels "democracy" normalized to a 0 to 1 scale. In this paper an alternative measure of political freedom and democracy is employed, using the Polity IV database.

[16] In the U.S. for example, the decline of inequality in the sixties and seventies corresponds with the institution of Johnson's "Great Society," while the subsequent rise in inequality in the eighties ocurred during the tenure of the more conservative Reagan administration.

[17] There is, of course, no good argument for ignoring relevant information, and data on income tax shares of GDP, social security expenditures and transfers is included in the analysis where it is available.

3 Data

To create a high-frequency, high quality data set to explore medium-run changes in inequality within countries, a 5-year panel was generated from the universe of existing observations and an extensive number of corrections and adjustments were made to the data. The basic methodology guiding the creation of the inequality data set used is outlined below. Further details are presented in Appendix A. Details regarding the set of covariates and their sources are given in Table 2, and some benchmark results using the Barro (1999) specification are presented in Table 3 to compare these various data sets.

3.1 Income inequality

A substantial effort was made to improve upon both the quantity and the quality of the standard data used in cross-country inequality research. The basic **DS** data set consists of an unbalanced panel of observations on Gini coefficients and quintile income shares. However, neither the quality (in both accuracy and comparability) nor coverage of the **DS** data set reaches the standards of most international macroeconomic data sets The original **DS** data set of "high quality" observations contains 108 countries with at least one Gini observation, 69 with two or more, 54 countries with four or more, and 32 with eight or more – representing less than 10% of the coverage of, for example, the Penn World Tables. More importantly, because this panel is unbalanced – so that countries have observations available in different years – the data require some manipulation to make meaningful statements about inequality in either a cross-section or a time series.

Traditionally, researchers have modified the **DS** set according to their individual needs and methodological viewpoints. To study the effects of inequality on growth in a panel data context, for instance, Forbes (2000) created a subset of six 5-year observations from 1966 to 1995, and only used countries with at least one set of contiguous observations (as required by her estimation technique). This gave her a sample of 45 countries, with 180 observations total. As a result of these limitations, her data set is very heavily biased towards richer countries (with better national economic data). Most of the developing countries included in Forbes' set only have one pair of observations. Barro (1999) uses a similar approach, but includes a number of additional observations for developing (largely African) countries which **DS** had not deemed "high quality" for lack of a primary source. Barro justifies his looser standard by arguing that the greater cross-sectional variation is more important than the potential increase in measurement error from using less than the highest quality data.

The approach taken in this paper follows that of Barro, making the first priority generating the widest range of data that can justifiably be called "acceptable." Then, to establish consistency across countries, years and survey types, two steps were taken. The first was interpolations across time to estimate "missing" data for specific years using observations from neighboring years, to make observations more comparable temporally by generating a panel across fixed five year intervals. Second, to control for variation in survey methodology across countries, corrections were made based on a series of hedonic regressions run on the universe of all available cross-country inequality data designed to estimate the bias imposed by measurement types.

The raw Gini coefficient data used in this dissertation were compiled from the World Income Inequality Database (**WIID**) available through the United Nations Development Program, which contains the **DS** data set as a subset of observations. Only **WIID** observations labelled "OK" were used. With a few exceptions, the observations used were also included among the "high quality" set of **DS** observations. Details regarding construction are discussed further in an appendix to this chapter.

The final inequality data set contains 514 income Gini coefficients, covering 137 countries from 1960-1995, at five year intervals. This is an unbalanced panel, covering 47% of a possible 1,096 observations. Although the richest countries have the most complete data, 70% of the observations are for countries classified by the World Bank as developing, a much higher share than is typical for work on income inequality. Of course, the total number of observations that can be employed varies depending on the availability of data on other variables of interest, and is typically much lower.

Table 3 compares estimates of the basic specification used by Barro (1999) to establish comparability among the various data samples used in this paper with that in previous work, and to assess the degree of robustness of the estimated relationships to the extended data. The first column presents the results reported by Barro. The second and third columns represent a similar specification using random effects estimator and data generated based on the method described in Barro (1999). The fourth and fifth columns present the same random effects estimates using the new inequality data set. The sixth and final column is the same as the fifth but using the adjusted Gini coefficient as a dependent variable, rather than including controls for survey type in the regression itself. The results suggest that the major characteristics of the data set are intact, but that the extended data set increases the number of observations substantially, lowering standard errors.

The variability in the parameter estimates across each column reflects two types of uncertainty facing the econometrician. The first arises from differences in choice of data sample, which is the basis for the standard parametric uncertainty accounted for in sampling distributions. Model specification, the source of differences between estimates reported in the second and third columns, is the second source of uncertainty. As suggested previously, however, choices with respect to data and model specification cannot be independent, since changes in sample composition occur when certain variables are included or excluded.

3.2 Covariate Data

To the best of its ability, economic theory must guide our choice of covariates, and the theoretical discussion of Section 2 has provided a long list of candidate variables, summarized in Table [ ] below.

[describe variables here]

4 The Methodology of Model Specification

A standard "textbook approach" to selecting an appropriate empirical specification involves an iterative a three-step process:[18] first specify an initial "best guess;" then use various diagnostics "tests" to evaluate the model; finally, change the model as needed, by respecifying the estimating equation, changing the procedure used, or transforming the data. The model is then re-estimated and the process repeated until the econometrician is satisfied the specification is "correct."

The classical approach to evaluating among competing models that can be nested is to use t and F tests, with the test statistic derived under the null hypothesis. Non-nested models present greater difficulty, however, since in the absence of a well specified null and alternative it is possible for test statistics to reject both models or neither model. Unlike hypothesis tests that evaluate between two hypotheses using one set of estimates, Hausman "specification" tests are tests of one hypothesis based on two sets of estimates. Rejection of a hypothesis is somewhat troubling, of course, in the absence a clear alternative.

Of course, any two non-nested models often can be thought of as subsets of a larger model, and this idea forms the basis of a literature on testing non-nested models.[19] The difficulty with diagnostics rooted in "testing down" from large initial models is that the dimension of the design matrix can be exceedingly large with respect to the size of the data sample. This, unfortunately, is often the case with cross-country regressions. As a result, the empirical growth literature has been characterized by the alternative of "testing up" from the basic Solow model. Testing-up involves the inferential hazards associated with non-nested tests, and is well known to increase the probability of making type I errors (including variables not actually in the "true" model).

In either case, a specification is generally deemed to be "correct" when the researcher

[18] Gujarati (1998) writes:

> After we obtain the results, we begin the postmortem....In determining model adequacy, we look at some broad features of the results such as the R^2 value, the estimated t ratios, the signs of the estimated coefficients in relation to their prior expectations, the Durbin-Watson statistic and the like. If these diagnostics are reasonably good, we proclaim that the chosen model is a fair representation of reality. By the same token, if the results do not look encouraging because the R^2 value is too low or because very few coefficients are statistically significant or have the correct signs or because the Durbin-Watson d is too low, then we begin to worry about model adequacy and look for remedies....

[19] Davidson and MacKinnon's J-test, in which the least-squares fitted values of one model are introduced as an additional regressor in a second model, and Ramsey's RESET test are examples of this type of nesting in which standard t and F-tests are used against the hypothesis that the coefficient(s) on the subset of variables comprising the alternative model equal zero.

cannot reject the absence of a) clear violations of the exogeneity restriction, $E(\varepsilon X) = 0$, based on the residuals, and b) omitted variable bias, as might be indicated by "incorrect signs" and/or a low R^2.[20] Whether pursuing a general-to-specific or specific-to-general methodology, selection of variables for final inclusion is typically based on hypothesis tests of statistical significance.[21] These hypothesis tests are specific to the null and alternative hypotheses, however, and will generally depend greatly on the other variables included in the specification used in the test.

In recognition of this, the literature has adopted the convention of reporting the results of several regression specifications employing different vectors of control variables as a way of providing evidence regarding the "sensitivity" of coefficients of interest to specification choice. If coefficient estimates are fairly stable across specifications, the results are deemed to be "robust" in the sense that we can be confident about the marginal effect independent of whether or not we consider the level of these additional control variables.[22]

In cross-country regressions, where a great deal of heterogeneity exists among the observations, a potentially large number of controlling variables is required to achieve exchangeability of the error terms. With multiple correlated controls in small samples, however, the geometry of least-squares will typically induce an additional (non-trivial) variance in coefficient estimates due to collinearity among the included controls. As a result, the conclusions drawn by authors employing this type of sensitivity analysis can suffer in one of three respects. First, for reasons of presentation clarity, the number of alternative specifications will generally be limited in number, and will therefore describe an incomplete set of potential biases. Second, despite the explicit uncertainty regarding specification, authors may choose to focus their conclusions on a particular "preferred" specification, using sensitivity analysis simply to argue that "similar results seem to hold" for other specifications. Although the strongest claims may reflect the most robust results, efforts are rarely made to capture this degree of (un)certainty with a formal probability statement.

Finally, the converse problem may arise if authors approach the variability in parameters

[20] Although omitted variable bias is an example of a violation of $E(\varepsilon X) = 0$, it is one that can remedied by correct variable selection, so the distinction is useful for later discussion. Other choices commonly thought of as specification, such as the estimator used, are often based on an efficiency criterion, an issue not addressed here.

[21] Specification of functional form is usually also incorporated within the hypothesis testing framework by testing restrictions placed on a (set of) parameter(s). These include the Ramsey RESET test for omission of higher order polynomial terms and tests on Box-Cox transformations of dependent variable.

[22] Suppose, for example, that the "true" distribution governing the data generation is $f(y|\theta)$, where y represents the data of interest, and θ an unknown parameterization of the distribution. Because of the complexity of the truth, we assume that attention is restricted to only a subset of the parameters. In particular, denoting the parameter set $\theta = (\alpha, \beta)$, assume attention centers on the parameters β only. The marginal distribution of β

$$p(\beta|y) = \int p(\beta|\alpha, y)p(\alpha|y)d\alpha$$

depends on α through the posterior distribution $p(\beta|\alpha, y)$. Sensitivity analyses therefore help to suggest the degree to which our beliefs $p(\beta|y)$ should be conditioned on assumptions regarding α.

across specifications by being overly qualitative – focusing, for example, only on the sign and statistical significance rather than specific values of their variables of interest, which would be more easily challenged. As an extreme case, Levine and Renalt (1992) offer an example of how a sensitivity analysis can lead one to the conclusion that "almost nothing is known for certain" about the determinants of economic growth despite hundreds of papers published on the topic.

Since the value of most empirical exercises lies in the ability to assess the quantitative effect of certain policies or variables, this hedging of results is unfortunate. While a proper appreciation for uncertainty in the analysis is as important as the estimates themselves, a more useful approach would be to incorporate a rigorous measure of our overall uncertainty (regarding both specification and parameters) in addition to point estimates. Leamer (1978) has shown that the task of inferring "expected" coefficient magnitudes from a range of estimates can be aided using Bayesian methods.

4.1 A Bayesian approach to model specification

In a Bayesian context, an econometric model can be thought of as the specification of a probability distribution over outcomes (or data, D) conditional on a set of assumptions, $\mathcal{A}$. This distribution is typically assumed to be parameterized by some vector θ, representing the quantity of interest for the econometrician. Econometric inference is based on the sampling properties of a given estimator $\hat{\theta}$ relative to some "true" value θ_0. Clearly, however, the "true" θ_0 depends on the "true" state of the world, given by $\mathcal{A}_0$. That is, the data generating process can be described as the distribution

$$p(D|\mathcal{A}_0) = \int pr(D|\theta_0, \mathcal{A}_0)p(\theta_0|\mathcal{A}_0)d\theta_0$$

Estimation by maximum likelihood is based on finding the mode $\hat{\theta}$ of the parameter distribution $p(\theta|D, \mathcal{A})$ given the model$\mathcal{A}$.[23] This section reviews a methodology of model selection based on searching for the mode of the distribution $p(\mathcal{A}|D, \hat{\theta})$. Model averaging techniques, discussed in the next section, are designed to incorporate information from the entire distribution $p(\mathcal{A}|D, \hat{\theta})$.

This probabilistic framework is useful in discussing the methodology guiding econometric choices made in empirical work, however the notion that there is some model $\mathcal{A}_0$ (with parameters θ_0) representing "the truth" that can be discovered by careful statistical analysis is not intended to be taken literally. A more appropriate characterization of the econometric process follows from the adage "all models are wrong, but some are useful." If all models are "wrong" then there is no point in attempting to find the "truth." Decisions regarding

[23] For simplicity, models will be referenced according to the set of assumptions $\mathcal{A}$ characterizing the probability distribution. Distinction between references to the set of hypotheses embodied in $\mathcal{A}$ and the subsequent probability model indexed by $\mathcal{A}$ should be possible from the particular context.

model choice therefore should be dictated by practical considerations and the ultimate goal of drawing inference correctly regarding some specific questions of interest.

As a result, perhaps, "specification" concerns in much econometric work typically involve addressing "estimation problems" involving violations of exogeneity restrictions or nonspherical residuals, checking the sensitivity of parameter estimates to the inclusion of additional controls, and ensuring the statistical significance in t-tests of the coefficients of interest. Not only is the R^2 rarely considered as a model diagnostic tool, its very use in such a context is widely considered to be inappropriate.[24]

However, the two goals of finding the "best" model fitting the data and generating appropriate inference on a specific empirical question are related. Faced with alternative sets of estimates from different model specifications, we should be more confident (all things equal) about the estimates from the model that is more likely to have actually generated the data. The question of which is the "most likely" model is conceptually complex, of course, and undoubtedly subjective. The probability a given model is correct (in the sense of it being the best approximation of reality among all models under consideration[25]) depends in part on reasoning and common sense, and in part on what the data reveal. Bayesian methods are once again helpful in formalizing and clarifying the issue, by suggesting the following proportionality relationship[26]

$$p(\mathcal{A}|D) \propto p(D|\mathcal{A})p(\mathcal{A}) \tag{3}$$

The left hand side represents the probability that model $\mathcal{A}$ generated the data observed, while the right hand side is the product of the data likelihood given the model $\mathcal{A}$ and the prior probability that model $\mathcal{A}$ is true. All evidence provided by the data are incorporated into the former, and all theoretical justifications for one model in terms of another (many of which are likely to be subjective) can be used to determine the latter. In this sense, no model known to be false can be "validated" by the data. However, absent strong *a priori* judgements, the

[24]The common lack of attention to the R^2 statistic may be a backlash against econometric work *too* focused on the R^2 in model evaluation because it is easily manipulated by including additional (extraneous) covariates and/or by transforming the dependent variables to reduce the size of the (unexplained) variance. Both of these are legitimate aspects of specification choice, however, and do not suggest that anything is "wrong" with the R^2 per se. Rather, it suggests that model choice should reflect a number of desirable properties (such as model parsimony and *a priori* sensibility) in addition to overall fit. Although the R^2 can be raised by adding covariates, excluding "insignificant" variables (typically those with t-statistics less than one) can raise the *adjusted* R^2 and other measures of model fit that penalize for the number of regressors. Certainly, the sequential learning process underlying final model specification choices should be guided by theory over characteristics of the data sample.

[25]Suggestions as to how the set of "all possible" candidate models should be derived in the face of our expressed uncertainty regarding the appropriate specification is taken up in Raferty (1995) and Draper (1995). The approach employed in the model averaging exercises of Section 7 begin from the set of all permutations of the variables suggested in the theoretical discussion of Section 2, and proceed according to the Occam's Window criteria suggested by Raferty (1995).

[26]The proportionality relationship $\propto$ is used in place of equality as we are ignoring the constant term representing $p(D)$. This constant, representing $\sum_M p(D|M)p(M)$, depends on the size of the model space and ensures that the distribution $p(\mathcal{A}|D)$ integrates to one.

posterior probability of a given model being true is determined by the likelihood of the data given the model, or

$$p(D|\mathcal{A}) = \int_\theta p(D|\theta, \mathcal{A})p(\theta|\mathcal{A})d\theta \tag{4}$$

The practical difficulty to applying the Bayesian methodology to estimate the model probability distribution 3 lies in calculating the integral in equation (4). A solution employed by George and McCulloch (1993), Smith and Kohn (1996), Madigan and York (1995), and Raferty, Madigan, and Hoeting (1997) uses Markov Chain Monte Carlo (MCMC) techniques to transition across the space of models. Rao and Tibshirani (1997) suggest bootstrapping the empirical distribution to approximate the posterior distribution. A less computationally intensive approximation can be achieved based on the Laplace method for integrals (Tierney and Kadane (1986)), reducing $p(D|\theta, \mathcal{A})$ to a function of the maximized likelihood $\mathcal{L}(\hat{\theta}|\mathcal{A}) \equiv \max_\theta p(D|\theta, \mathcal{A})$ and the Fisher information matrix based on the idea that, asymptotically, the mass of the probability density increasingly lies within a small neighborhood of the maximum likelihood estimator.

This offers some justification for the use of the R^2 as a diagnostic tool in model selection, since the maximized likelihood is directly related to the R^2 for an estimated model $\mathcal{A}$ with normally distributed residuals[27]. However, the (unadjusted) R^2 fails to account for the size of the parameter vector, and therefore cannot address the optimal dimensionality of the model. Two alternative measures of "fit" suggested by likelihood theory are the Akaike Information Criterion (AIC)

$$AIC = \mathcal{L}(\hat{\theta}) - 2q$$

and the Bayesian Information Criterion (BIC)[28]

$$BIC = \log \mathcal{L}(\hat{\theta}) - \frac{1}{2} q \log(n)$$

for a model with q parameters, given n observations. The AIC was derived by Akaike (1973) as an estimate of the (relative) expected Kullback-Leibler distance between a given approximating model and the unknown truth.[29] The BIC, derived by Schwartz (1978) within

[27] Both the maximized log-likelihood and R^2 are decreasing in the residual sum-of-squares. In particular, for a constant C,

$$1 - R^2 = \frac{C}{\mathcal{L}(\hat{\theta})^{2/n}}$$

[28] The AIC and BIC are also presented as $-2\log\mathcal{L}(\hat{\theta}) + 2q$ and $-2\log\mathcal{L}(\hat{\theta}) + q\log(n)$, respectively. Obviously, these transformations do not affect any of the arguments made regarding model selection.

[29] The Kullback-Leibler distance (KLD) between a probability distribution given by $p(x|\theta)$ and the (unknown) true process, denoted $f(x)$ is given by

$$\begin{aligned} & \int f(x) \log\left(\frac{f(x)}{p(x|\theta)}\right) dx \\ = & \int f(x)\log f(x) dx - \int f(x) \log p(x|\theta) dx \end{aligned}$$

a Bayesian framework, reflects an asymptotic approximation of the posterior probability that a given model is the "correct" one from among a set of candidates. Practically speaking, the BIC tends to penalize larger models relative to the AIC; within their individual contexts, however, each of these measures can be thought of as representing the relative likelihood of the model given the data.

Evidence for a model $\mathcal{A}_1$ relative to an alternative $\mathcal{A}_0$ can be summarized by the posterior odds ratio $OR_{1,0} \equiv p(\mathcal{A}_1|D)/p(\mathcal{A}_0|D)$, which by Bayes' rule is equal to

$$OR_{1,0} = \frac{p(D|\mathcal{A}_1)}{p(D|\mathcal{A}_0)} \frac{p(\mathcal{A}_1)}{p(\mathcal{A}_0)}$$

This is the Bayes factor times the ratio of model priors. If the priors across all models are treated as being equal, then Raferty (1995) suggests a model selection scheme based on the BIC approximation to the posterior density, employing the notion that evidence favoring one model relative to another is simply $\log OR_{1,0} \approx BIC_1 - BIC_0$. Normalizing to ensure that the posterior model probabilities sum to one, Rafferty uses the BIC for estimated model $\mathcal{A}_m$ as an estimate of $p(\mathcal{A}_m|D)$.[30]

4.2 Bayesian Inference and Model Averaging

In the course of the model selection, many models with roughly similar likelihoods are encountered. This naturally leads one to ask *how much better* is the "best" model? An alternative to using the posterior mode estimate of θ – or the "most likely value of the most likely model" – is to use the posterior mean. Over the set of models (indexed by m), the posterior mean estimate of θ represents an average value of θ across the space of models, to provide a "specification-robust" estimate of the distribution of θ based on the law of total probability

$$p(\theta|D) = \sum\nolimits_m p(\theta|D, \mathcal{A}_m) p(D|\mathcal{A}_m)$$

Two major issues face the econometrician employing model averaging techniques: calculation of the posterior density weights $p(\mathcal{A}|D)$ discussed in the previous section, and the choice of the model space. Different approaches have been suggested with respect to the latter.

Draper (1995) proposes taking the a priori candidate model, and then averaging in a "neighborhood". This approach works well when a basic specification is clear and the issue is the set of appropriate controls. When theory is more ambiguous and the appropriate a priori

The first term integrates to (an unknown) constant, which implies that estimates of the KLD can only be evaluated relative to one other, not on an absolute scale. The second term represents the quantity $E_x \log p(x|\theta)$. This expectation requires knowledge of the true data process $f(x)$ and the parameter vector θ. The AIC was therefore derived as an estimate of minus two times the *expected* distance, given the data D, or $E_D E_x \left[\log(p(x|\hat{\theta}(D)\right]$. Although the AIC does not vary with sample size as does the BIC, Burnham and Anderson (1998) suggest adding a bias correction term $\frac{2K(K+1)}{n-K-1}$ to the AIC in small samples.

[30] It should be noted that the BIC approximation to the posterior probability is made without reference to any particular (parameter) prior, nor does does it require one.

model is less clear, however, some rule is necessary to sort through the set of potential models. This set can be quite large, as the permutations of a possible candidate variable set grow exponentially with the dimension of the design matrix. Rafferty proposes an algorithm for sorting through the set of candidate models using an approach he terms "Occam's Window". This approach is characterized by two rules:

1) Accept a model when it has a preponderance of the evidence; and

2) If a model is rejected, then all subsets of the model are rejected.

This pares the model space to a more easily managed set. Posterior densities are approximated using the BIC approximation.

4.3 Model specification as "variable choice"

The process of model selection embodies a number of choices that must be made by the econometrician to ensure an appropriate conditioning process. In addition to uncertainty regarding the "correct" set of covariates to include and their functional form, the econometrician faces uncertainty regarding the distribution of parameters and residuals. In particular, when substantial heterogeneity exists among the units of observations, heterogeneity among the estimated parameters and/or the residuals may result. That is, there is uncertainty as to whether the observations should be treated as draws from the same model. Finally, there are the usual questions of whether corrections for correlation among the errors and/or between the errors and covariates are necessary.

To a large extent, concerns regarding both functional form and parameter heterogeneity can be incorporated within the framework of variable selection.[31] In particular, unknown functional forms can be approximated by the inclusion of higher order powers and interactions of variables appearing in the candidate set, while parameter heterogeneity can be incorporated by including interactions with dummies signifying the appropriate distinguishing characteristics (e.g. country or regional identity). Of course, what we can hope to learn from the data depends to a great deal on what restrictions we are willing to impose, and the introduction of additional flexibility will have the usual cost of decreased precision of the estimates. The amount of flexibility desired can be weighed against the costs of the relaxed structure, and choices can be made for each variable ranging from straight linear (first order) effects to non-parametric curve fitting by natural cubic splines.

Similarly, some uncertainty regarding the error structure can also be handled within our

[31] One exception includes transformations of the dependent variable. Because the Gini bounded between zero and one, a transformation (e.g. $\ln(Gini/1 - Gini)$) might seem appropriate, since the projection $X\beta$ can conceivably take on any range of values on the real line. However, while the theoretical range of the Gini coefficient is [0,1], the support of the data is roughly .20 to .60, suggesting that censoring is not likely to be a problem in estimation. Over this support, the transform $\ln(Gini/1 - Gini)$ is roughly linear in $X\beta$, although the interpretation of the coefficients does change since the slope of the transformation function does not equal one. In keeping with the literature to date, and to preserve simple coefficient interpretations as marginal changes in the Gini coefficient, no tranformation was used.

variable selection framework. While careful post-estimation diagnostic testing remains the best approach to assessing whether "econometric fixes" are necessary for serial correlation and exogeneity violations, heteroskedasticity – and in particular, the identification of outliers – can also be incorporated into the Bayesian framework of variable selection. The observed data can modeled as draws from a mixture of the principal model distribution and a second distribution of "outliers" characterized by a higher variance. Given some prior specification on the mixture, the selection of "outliers" (and decisions regarding the weight it should receive in regression fitting) can then be assessed using the Bayesian posterior probability that a given observation is from the distribution with higher variance.

4.4 Some Practical Challenges to Specification Search

Data-driven variable selection techniques suffer from both practical and methodological problems. Practically, stepwise variable selection or model "pre-testing" can overemphasize the importance of random features of the data and generally complicate standard hypothesis testing. Methodologically, the idea that both parameters of interest and the structure of the relationships among them can be inferred correctly from only the information available in the sample data seems unlikely; that specification can or should take place in isolation from any intuitive logic also seems troubling.

Bayesian methods have been developed principally as a way to introduce formally prior into sample information. However, it is not clear that all available information can be specified probabilitstically – there is almost always a process of learning undertaken by the researcher in analyzing the data that is difficult to quantify. This section details areas in which practical choices were required in order to implement the Bayesian variable choice methodology. These include how to maximize the initial set of available observations and which variables to use to test particular theories. The three criterion employed were the theoretical justification, the resulting observations available for regression, and the avoidance of multicollinearity. Endogeneity, another concern, is discussed in the final section discussing time series variation.

1) Not enough observations Because cross-country data sets are invariably incomplete, the larger the set of covariates included in a regression, the smaller the number of observations available will be. In the context of variable selection searches, where a large set of candidate variables are being considered, the problem is particularly severe as the intersection of observations with no missing data can quickly go to zero. Two strategies were employed to maximize the set of available observations. The first was to make available observations a criterion when choosing a particular proxy from among a set of variables measuring a given theoretical quantity of interest. The second was to impute missing data using various conditioning schemes, discussed in Appendix B.

	Full Sample 509	Sample 381	Sample 279	Sample 107
Mean	39.5	41.0	43.0	39.2
Std. Dev	11.0	10.4	9.7	9.5
(within)	3.9	3.4	3.4	2.3
(between)	11.0	10.4	9.8	10.3
Minimum	17.3	21.1	21.8	21.8
Maximum	67.8	67.8	67.8	67.4
# countries	135	95	84	49
average T	3.8	4.0	3.3	2.2
countries w/				
T > 6	21	18	4	0
T > 3	65	50	37	15
T = 1	25	17	18	18

Sample 509: Gini coefficients only (no covariates)
Sample 381: includes covariates PCGDPSH through AGSHARE
Sample 279: also includes INDSHARE through AGVAPC
Sample 107: also includes BIRTHRATE through CATHO80

Figure 3: Summary statistics of the Gini coefficient over four samples of decreasing size, representing the total observations available for regression with an increasing set of covariates.

2) Proxy searches In many cases, data on several variables exist that capture essentially the same economic forces. For instance, sectoral employment shares and shares of value added both capture aspects of the Kuznets-type structural adjustment process, while the real interest rate, inflation, lending rate, and deposit rate each capture elements of both the cost of capital and the opportunity cost of education. Where data on shares exists – for sectoral and educational data, for instance – decisions must be made regarding how to summarize the distributional information into a one-dimensional summary for regression analysis. The particulars of these choices are also discussed in Appendix B.

3) Distinguishing among correlated variables. Variable selection and model choice decisions are particularly complicated when correlations among the set of potential regressors are high. It makes evaluation of and selection between alternative models difficult. This can be problematic when those models have widely different implications for inference. For instance, if oil-exporting, Middle-eastern, Islamic, middle-income countries are found to have particularly high inequality relative to other countries, is that a result of trade, regional history, institutions, or income level? Inference is accordingly complicated when one cannot distinguish from the sparse empirical evidence which variable(s) are causal and which are simply correlated with the causal variables. It is possible

Leamer (1978, p. 171) discusses how this issue of multicollinearity can be thought of as a special case of the "weak evidence problem" in which the particular challenge is assigning contributions by individual regressors among the set of collinear variables.[32] The fact that problems of multicollinearity diminish asymptotically provides little consolation in cross-country regressions where the sample size is fixed at a relatively small number. Leamer (1978, p.173-181) shows how in a Bayesian analysis, estimates are well specified given a particular prior, but collinearity generates an extreme sensitivity of posterior distributions to the prior distribution.[33] Leamer suggests a measure $c_i \geq 1$ that indexes the degree to which existence of other parameters causes problems for interpretation problems for coefficient i. He then states that ranking coefficients by $t-$statistics is equivalent to an inverse ranking of coefficients by c (Leamer, 1978, p. 177, n. 18).

To sort out correlations, variance in the Gini coefficient was decomposed among several factors (some omissions were necessary to prevent perfect collinearity in the sample). The results, presented in Figure 4, suggest that geography (specifically, being in the tropics), legal systems, and income play a more significant role in explaining inequality than the type of exports. The impact of being a manufactures exporter is still large and negative, but compared to a regression on export type alone, it is no longer significant and substantially lower (without controlling for other characteristics, export manufacturing countries have Gini coefficients 10 points lower on average). As a result, export composition was excluded among our list of candidate covariates, to focus attention on income, geography, and the legal system.

5 Evaluating the Existing Literature

As an initial step to uncovering the determinants of inequality, three specifications described earlier – those of LSZ, GJ, and Barro – were nested into a single model and estimated using the set of available observations (106 in total). Based on F-tests, none of the three models could be rejected wholesale. The next step was to evaluate the variables in each specification individually. First, an iterative forward and backward search was employed to maximize the AIC. The final resulting model differed according to whether region and time (decade) fixed effects were allowed in the initial specification. The evidence is presented in Figure 5.

A Kuznets relationship appears in both sets of estimates, as does the role of primary

[32] Goldberger (199?, p. ?), and Leamer (1978, p.171).

[33] "Thus the collinearity problem is transformed from a problem of characterizing and interpreting a multidimensional likelihood function into a problem of characterizing and interpreting a multidimensional prior distribution." (Leamer, 1978, p.176). Like a true Bayesian, however, Leamer leaves the question of how one should formulate these prior distributions to the individual.

Analysis of Variance:	**mean group effects**		
	Coefficient	Std. Error	t-stat
Constant term	42.0	4.3	9.9
Income level			
Low	-0.4	0.9	-0.5
Upper-middle	2.1	1.0	2.2
OECD	-7.3	1.1	-6.5
Tropical country dummy	4.9	0.9	5.4
Legal system			
French	2.6	0.8	3.3
Socialist	-12.5	1.1	-10.9
Scandanavian	-1.9	1.5	-1.2
Export type			
manufactures	-1.9	4.1	-0.5
primary (non-fuel)	0.6	4.1	0.2
fuel (mainly oil)	-0.9	4.2	-0.2
services	-0.1	4.2	0.0
diversified exporter	0.1	4.1	0.0
Decade			
1970	-0.4	0.8	-0.6
1980	-1.8	0.8	-2.3
1990	1.1	0.8	1.4
483 obs, pooled	R^2 = .59		

Figure 4:

education in reducing inequality. Secondary education remains in the fixed effects case.[34] The use of regional effects also leads to the index of democracy remaining in the final specification. For theoretical reasons, interactions with the level of income were allowed for both democracy and trade, even though they did not appear in the original authors' specifications. In the former case, the nature and effectiveness of democratic institutions in providing redistributive pressure are allowed to change with the level of development, while in the latter case, the effect of international trade on inequality is assumed to depend on countries' endowments of labor and capital, proxied by GDP/capita.

Next, standard hypothesis testing was used to "test down" from an enlarged specification of the nested model, which included higher order powers for the industrial share, inflation, and private credit, and the separation of between- and within- country effects of per-capita income levels. Using even the relatively high benchmark p-value of 30%, the null hypothesis of zero could not be rejected for a number of variables including both the land Gini coefficient and private credit, providing some evidence against the importance of credit constraints in explaining the level inequality. This stands in marked contrast to the results of Li, Squire, and Zou (1998).

[34] Average years of higher education appear as well when regional, but not time dummies are allowed.

Results of Stepwise Variable Selection

Covariates	w/ region & & time effects coefficient	t - stat	no region or time effects coefficient	t - stat
lny	101.3	5.1	81.5	5.5
lny2	-5.7	-4.6	-4.5	-4.9
pyr	-1.1	-2.5	-0.9	-2.1
syr	1.2	1.8		
indshare	-0.67	-7.3	-0.67	-7.1
under15	36.0	2.4	61.5	4.2
over65	51.4	1.9	76.2	2.7
inflation			0.05	2.6
transfers	0.07	1.5	0.08	1.6
trade	0.3	1.4	0.5	2.4
trade:lny	-0.04	-1.6	-0.06	-2.6
landgini	0.04	1.5	0.07	2.4
credit	0.03	1.8	0.04	2.1
dma	1.85	1.7		
dma:lny	-0.23	-1.8		
reg.lac	3.7	2.7		
reg.ssa	19.2	5.5		
(Intercept)	-405.0	-4.9	-337.6	-5.3
R^2	0.76		0.70	

Figure 5: Final model based on backwards elimation from a full model nesting the specifications of Barro (1999), Gustaffson and Johansson (1999), and Li, et. al. (1998).

As mentioned earlier, stepwise variable selection procedures based on the maximization of "fit" can lead to specifications that are overly sensitive to characteristics of the particular data sample. To evaluate more carefully how robust the support for these variables under alternative specifications, Rafferty's BIC approximation to the model posterior was used to assist in estimating model averaged coefficient estimates. The results, shown in Figure 7, suggest that the Kuznets curve appears as one of the more robust relationships in the data, along with primary school attainment and the share of employment in industry. When dummies for Latin America and Sub-Saharan Africa are included, they also appear to have strong and robust effects across a range of models, suggesting the importance of some omitted variables in these regions. When these region effects are excluded, the rate of inflation and the land Gini coefficient – both of which are particularly high in these regions – become significant correlates of inequality. Based on the posterior mean estimates, the estimated Kuznets curves appear to be particularly steep in ascending at low levels of income, falling only gradually as incomes rise further. This suggests that economic growth may not be a complete remedy to alleviating income inequality in the long-run.

The Results of "Testing Down" from a Broad Specification
based on variables suggested by previous empirical literature

Covariate	Coefficient	Std. Error	t-statistic	Mean	Std. Dev.	coefficient magnitude*
	(num. obs =169)			(Summary Statistics)		
mlny	49.43	23.02	2.15	8.1	0.9	**4.23**
mlny2	-2.45	1.44	-1.70	65.9	15.0	**-3.34**
wlny	4.06	2.83	1.43	0.0	0.3	0.11
pyr	-1.79	0.48	-3.73	3.7	1.7	-0.28
syr	1.03	0.63	1.64	1.4	1.1	0.11
indshare	-0.90	0.30	-3.04	28.8	14.0	**-1.16**
indshare2	0.01	0.00	1.90	1023.2	823.8	0.62
under15	0.22	0.16	1.39	33.4	10.0	0.20
over65	0.47	0.28	1.69	7.1	4.2	0.18
inflation	-0.12	0.05	-2.33	16.3	26.3	-0.28
inflation2	0.00	0.00	2.67	954.8	3712.1	0.37
trade	0.44	0.18	2.39	64.7	49.6	**1.98**
trade_lny	-0.05	0.02	-2.51	529.3	457.2	**-2.23**
transfers	0.05	0.04	1.07	37.7	21.3	0.09
dma	3.70	1.07	3.45	2.6	7.3	**2.48**
dma_lny	-0.47	0.13	-3.52	26.3	60.8	**-2.63**
t7	-3.22	1.43	-2.26	0.23	0.42	-0.12
t8	-1.38	0.92	-1.50	0.31	0.46	-0.06
reg_ssa	22.25	3.61	6.17	0.14	0.35	0.70
reg_lac	10.37	1.98	5.23	0.19	0.40	0.38
reg_eap	5.92	1.43	4.13	0.16	0.37	0.20
reg_eca	11.39	2.93	3.89	0.20	0.40	0.41
reg_mena	5.54	1.80	3.08	0.06	0.24	0.12
constant	-187.02	92.53	-2.02			

removed (for p-values greater than 30%)

credit	p =	92%
hyr	p =	86%
reg_sa	p =	88%
landgini	p =	79%
govc	p =	72%
reg_we	p =	58%
wlny2	p =	50%
credit2	p =	35%

* coefficient magnitude represents the impact of a one standard deviation shock in the independent variable in terms of the standard deviations of the Gini coefficient. Bold face represents a greater than one-for-one standard deviation impact.

Figure 6: Results of the general to specific methodology.

BAYESIAN POSTERIOR MODEL PROBABILITIES & POSTERIOR MEAN ESTIMATES

Specification 1: covariates from Barro, GJ, & LSZ (region dummies for Sub-Saharan African and Latin America included)

MODEL (n=163)				COVARIATES									
Rank	R^2	Post. Prob.		lny	lny2	pyr	hyr	indshare	under15	trade	trade*lny	reg.ssa	reg.lac
1	72.1	18%		x	x	x		x				x	x
2	73.0	16%		x	x	x	x	x				x	x
3	73.6	10%		x	x	x		x		x	x	x	x
4	72.7	9%		x	x	x		x		x		x	x
5	73.5	7%		x	x	x	x	x		x		x	x
6	72.5	4%		x	x	x		x	x			x	x
total prob.		63%											
next 34		37%											
posteriors			β	88.0	-5.0	-1.48	1.81	-0.55	1.45	0.04	-0.01	19.74	5.31
			s.e.	14.4	0.9	0.51	2.82	0.08	5.11	0.14	0.02	3.28	1.17
			Pr(not 0)	*100*	*100*	*100*	*36.1*	*100*	*11.3*	*29.3*	*40.4*	*100*	*100*

Specification 2: covariates from Barro, GJ, & LSZ (region dummies excluded, but with time dummies)

MODEL (n=163)				COVARIATES										
Rank	R^2	Post. Prob.		lny	lny2	pyr	indshare	under15	over65	inflation	credit	landgini	t9	gc
1	67.0	14%		x	x	x	x	x		x		x		
2	65.8	8%		x	x	x	x	x		x				
3	67.7	6%		x	x	x	x	x		x	x	x		
4	67.7	5%		x	x	x	x	x		x		x	x	
5	66.6	5%		x	x	x	x	x		x			x	
6	65.4	4%		x	x	x	x	x				x		
7	68.5	3%		x	x	x	x	x	x	x	x	x		
8	67.4	3%		x	x	x	x	x	x	x		x		
total prob.		49%												
next 34		51%												
posteriors			β	54.7	-2.9	-1.32	-0.52	42.8	18.2	0.06	0.01	0.05	0.56	-0.02
			s.e.	14.3	0.9	0.51	0.09	17.2	32.6	0.03	0.02	0.04	0.99	0.06
			Pr(not 0)	*100*	*100*	*96.3*	*100*	*99.2*	*30.5*	*93.6*	*24.3*	*65.1*	*30.6*	*11.3*

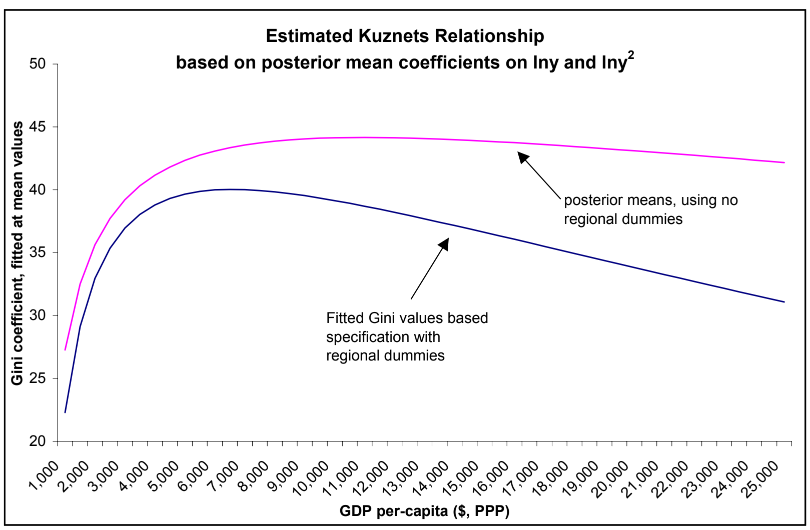


Figure 7: Posterior probabilities over the set models encompassing regressors employed in previous empirical work, and the estimated Kuznets relationship.

6 The evolution of inequality within countries

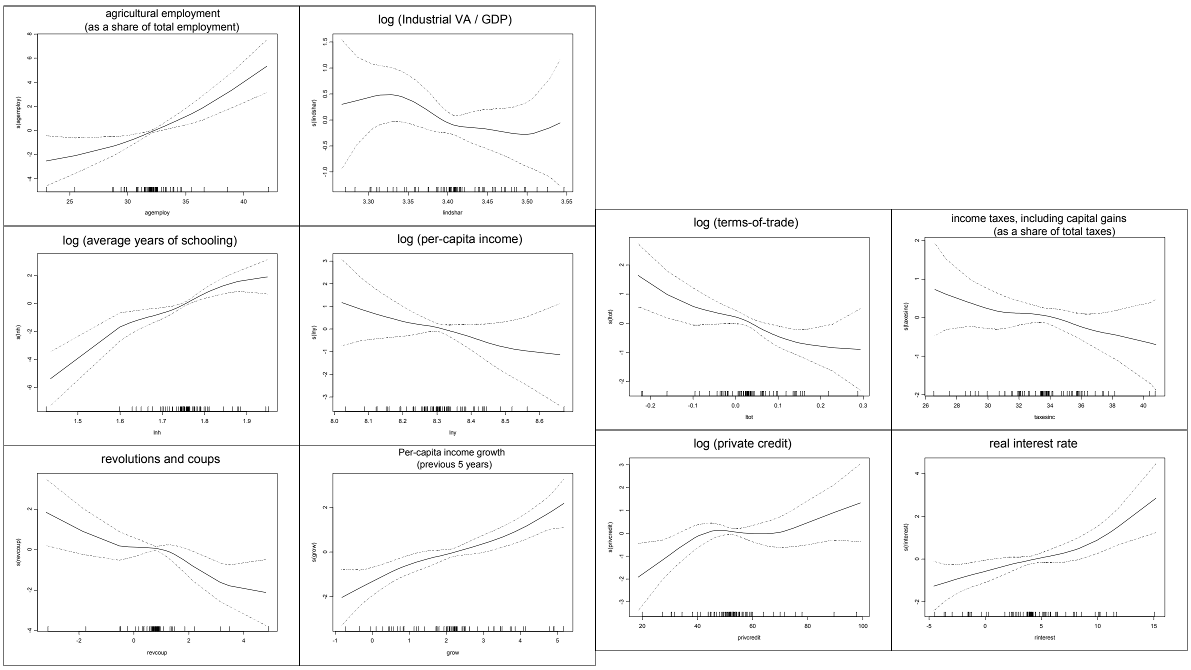
The determinants of inequality within countries over time (in the short to medium run) may differ from the determinants of inequality across countries (reflecting the long-run). This section focuses on inequality dynamics within countries by implementing a country fixed effects estimator to absorb cross-country variation in average levels of inequality. This allows us to isolate and focus specifically on the variables suggested in the theory section that might affect inequality over the short to medium run. These include, potentially, the Kuznets curve relationship (with respect to both sectoral employment and income), the rate of growth, and changes in the terms of trade. In addition, changes in private credit, the real interest rate, and political shocks and events (such as revolutions and coups) that might breed social instability in the country were examined.

Again, the model search began with a general to specific methodology, testing down from a "full" model or 13 covariates plus an intercept. Despite an attempt to consolidate variables in order to minimize model size, the presence of country fixed effects left relatively few degrees of freedom (72 observations were left after differencing to absorb the country dummies). When using a cubic smooth for agricultural employment (rather than a simply quadratic polynomial), the model dropped only the "income tax" and "assassination" variables. To avoid possible endogeneity problems (discussed in Appendix B), the model was run without the assassinations, revolutions & coups, and government crises variables, while introducing quadratic terms on all variables to assist in flexibility. The final model chosen is presented in the table below.

Variable	S.D. (σ)	Coefficient (β)	t - stat	$\Delta G (= \beta\sigma)$
intercept	–	25.15	2.42	
agemploy^2	176.4	0.007	5.20	1.23
growth	1.22	0.469	4.76	0.57
ln(h)	0.07	11.99	5.84	0.84
ln(y)^2	1.71	-0.22	-1.55	-0.38
ln(ToT)	0.084	-6.37	-5.01	-0.54
priv. credit	12.16	0.038	2.98	0.46
rinterest	3.83	0.205	5.66	0.76
$obs = 72$	$R^2 = .71$	$F_{7,64} = 22.8$		

6.1 Estimating functional forms

The quadratic terms of both the agricultural share and log income appear in the final model but not the principal effects. This is due to the penalty imposed in the AIC on unnecessary regressors. A more flexible additive non-parametric model was estimated using the set of regressors found to be significant in the parametric regression to better approximate the functional relationships.[35] The results appear in the Figure below. The only instances in which linearity can be clearly rejected are in industrial value added as a share of GDP, private credit, and the (log) of average years of schooling. The quadratic approximation of the agricultural share found in the parametric regressions is apparent again, although the relationship could fairly easily be described using a simple linear approximation.



The non-linearity in the effect of private credit is of some interest. While conventional theory holds that initial rises in private credit will alleviate credit constraints faced by the poor, facilitating their investment, quite the opposite seems to take place. Increases in private credit are initially associated with rapid increases in inequality, after which the impact seems to disappear. Although the graph seems to imply it rises again later, the data in this region is probably too scarce to draw a firm conclusion. Finally, it is interesting to note that as theorized in the introduction, positive changes in terms-of-trade are associated with marked declines in income inequality. In theory, the connection between international trade and factor incomes works entirely through the terms-of-trade (i.e. relative goods prices of exports relative to imports) but the focus of most linkages of this nature have relied on quantitative measures such as the volume of trade.

[35] The estimation of such generalized additive models, described in Hastie and Tibshirani (1990), involves an iterative "leave-one-out" process in which, for each covariate, the residuals from a fit using the remaining variables are smoothed. This smooth is then used to generate the residuals for the next covariate to be smoothed.

BAYESIAN MODEL ASSESSMENT BASED ON POSTERIOR PROBABILITY

Model Ranking	posterior probability	BIC	variables included							
1	13.1%	-58.6	agemploy	grow	lnh	ltot	privcredit	rinterest		
2	8.7%	-57.8	agemploy	grow	lnh	ltot	revcoup	rinterest	taxesinc	
3	6.0%	-57.0	agemploy	govtcrises	grow	lnh	ltot	privcredit	rinterest	
4	5.5%	-56.9	agemploy	grow	lnh	ltot	privcredit	revcoup	rinterest	taxesinc
5	5.4%	-56.8	agemploy	grow	lnh	lny	ltot	privcredit	rinterest	
6	5.2%	-56.7	agemploy	grow	lnh	ltot	privcredit	rinterest	taxesinc	
7	5.1%	-56.7	agemploy	grow	lnh	lny	ltot	privcredit	revcoup	rinterest
8	4.9%	-56.6	agemploy	grow	lnh	ltot	privcredit	revcoup	rinterest	
9	3.8%	-56.1	agemploy	grow	lnh	ltot	rinterest			
10	3.6%	-56.0	agemploy	grow	lnh	ltot	rinterest	taxesinc		
total	61.3%									
next 25	38.7%									

Posterior model probabilities calculated by BIC approximation (Rafferty, 1994)

Total probability that coefficient is non-zero and posterior estimates

Variable	Probability	Coefficient	Std. Error
agemploy	100.0	0.50	0.09
$agemploy^2$	2.8	0.00	0.00
assass	1.6	0.00	0.01
govtcrises	24.7	-0.06	0.14
grow	100.0	0.51	0.11
lindshar	13.3	-0.41	1.44
lnh	100.0	11.72	2.28
lny	25.7	-1.18	2.44
ltot	100.0	-6.17	1.33
privcredit	75.5	0.02	0.02
revcoup	44.5	-0.11	0.16
rinterest	100.0	0.20	0.04
taxesinc	38.8	-0.03	0.05
(coefficient)		14.45	22.52

Figure 8: The determinants of within-country variation: posterior model probabilities for models including country fixed effects.

6.2 Bayesian Model Averaging

Results for the fixed effects ("within") estimator are shown in Figure 8. In this case, the Occam's Window generate a "preferred" model does not include log income and the industrial share, nor income taxes and the governmental event variables. It should be noted, of course, that although the probability of our preferred model is substantially larger than that of the next candidate model, there is an 87% probability that it is not, in fact, the correct model. The posterior model averages of the coefficient estimates and standard deviations suggest that changes in the level of agricultural employment, economic growth, years of schooling, terms of trade, and the real interest rate almost surely play a role in corresponding changes in income inequality. Private credit likely does as well, but the model-weighted probability that the remaining coefficients are not equal to zero falls off quickly.

7 Conclusion

The economic literature on the cross-country income distribution entails a vast array of theories (such as international trade, technology, property rights, credit markets, political economy, etc.) and mechanisms (including factor prices, accumulation decisions, and government redistribution), with little evidence on the role that most of these play in the observed pattern of income inequality across countries. Despite recent improvements in the availability of cross-country inequality data, several pitfalls exist for cross-country empirical research on inequality. First, identification of the effects of individual variables is complicated by the lack of clear theoretical relationships between these variables and the income Gini coefficient to guide specification in addition to the collinearity among many of these candidate explanatory variables. Second, the data itself is incomplete and often of questionable accuracy. While one approach to dealing with the former issue of specification uncertainty is to estimate large, flexible models, the constraints posed by the latter present a clear practical constraint.

Following a recent line of research employing a Bayesian probabilistic approach to model specifications, this paper reviews and evaluates several competing, non-nested empirical models of the Gini coefficients suggested in the literature to date. Model probabilities can be used to gauge the coefficient robustness in a way that conventional sensitivity analyses fail to summarize. Our qualitative understanding of the processes affecting inequality is aided by calculating the posterior probabilities that coefficients are not equal to zero. Excluding variables from specifications that do not appear to be significant based on their estimated posterior densities can be justified using standard arguments of the bias-variance tradeoff. Model averaging thus provides an approach to "model uncertainty" that is both more intuitive and more practical than standard hypothesis testing.

The results of a general to specific (or, "nest it, test it") methodology and model averaging both suggest that the Kuznets curve – an inverse U-shaped relationship between the Gini coefficient and per-capita income – remains one of the more robust relationships in a cross-section of countries, after controlling for a range of variables. Within countries, however, economic growth appears to have a strong, positive, and largely linear relationship with inequality. Inequality tends to be higher on average in periods proceeded by more rapid economic growth, and to rise over time in conjunction with increases in per-capita income.

The role of education also appears to be significant, although the importance of years of higher education found by Barro does not appear robust to alternative model specifications. Rather, the number of years of primary schooling (which can be thought of as simply proxying for the share with no education) appears to be the most robust determinant of income inequality. This provides an additional justification for arguments that have been made in the development community for prioritizing access to primary education among various policy objectives.

Interestingly, the impact on inequality of increases in private credit, a variable often used to proxy for (the absence of) credit market imperfections, appears to be positive as is the real interest rate. This stands in contrast to previous results of Li, Squire, and Zou (1998), who find a negative coefficient on the level of M2/GDP, another measure of financial market development, and suggests that increasing the size of credit markets may favor relatively more wealthy investors rather than increasing the opportunities of the poor. The share of trade in the economy also appears to be to raise inequality, but at a decreasing rate as per-capita income rises. Although this seemingly runs counter to the conventional wisdom born out of neoclassical trade theory, it is consistent with the idea that the institutions of more developed nations are better able to counter the volatility created by international integration. Within countries, a rise in the terms of trade appears to be associated with falling inequality.

The data-driven approach to specification adopted in this paper is intended as a starting point for investigation of the data, and is justified by the lack of a clear theoretical framework to guide empirical analysis and the need for a formalistic approach to evaluating between competing specifications presented in the literature. Bayesian estimation of the posterior density of the set of potential covariates conditional on the data, but not a particular model, helps generate a more accurate assessment of the sampling distribution of these variables in the face of model uncertainty. Ultimately, however, using theory to place additional *a priori* structure on the data will improve the accuracy of the estimates as well as our understanding of the appropriate specification.

References

Acemoglu, D. (2002): "Technical Change, Inequality, and the Labor Market„" *Journal of Economic Literature*, XL, 7–72.

Bannerjee, A., and E. Duflo (1999): "Inequality and Growth: What Can the Data Say?," MIT.

Barro, R. J. (1991): "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics*, 106, 407–443.

——— (1999): "Inequality, Growth, and Investment," NBER Working Paper No. W7038.

Barro, R. J., and J.-W. Lee (2000): "International Data on Educational Attainment Updates and Implications," NBER Working Paper No. W7911.

Bénabou, R. (1996): "Inequality and Growth," NBER Working Paper 5658.

Brock, W. A., and S. N. Durlauf (2001): "Growth Economics and Reality," *World Bank Economic Review*, 15 (2), 229–272.

Burnham, K. P., and D. R. Anderson (1998): *Model Selection and Inference*. Springer, New York.

Cowell, F. A., and S. P. Jenkins (1995): "How Much Inequality Can We Explain? A Methodology and Application to the USA," *The Economic Journal*, 105, 421–430.

Deininger, K., and L. Squire (1996): "A New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10(3), 565–91.

Draper, D. (1995): "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society (B)*, 57, 45–70.

Easterly, W., and M. Sewadeh (2001): "Global Development Network Growth Database," The World Bank (www.worldbank.org).

Feenstra, R. C., and G. H. Hanson (1996): "Foreign Investment, Outsourcing, and Relative Wages," in *The political economy of trade policy: Papers in honor of Jagdish Bhagwati*, ed. by R. C. Feenstra, G. M. Grossman, and D. A. Irwin, pp. 89–127. MIT Press.

Feldstein, M. (1999): "Reducing poverty, not inequality," *Public Interest*, (137).

Fernández, C., E. Ley, and M. J. F. Steel (2001): "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 16(5), 563–576.

Forbes, K. J. (2000): "A Reassessment of the Relationship Between Inequality and Growth," *American Economic Review*, 90, 869–887.

George, E. I., and R. McCulloch (1993): "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

Greenwood, J., and B. Jovanovic (1990): "Financial Development, Growth, and the Distribution of Income," *Journal of Political Economy*, 98(5), 1076–1107.

Gujarati, D. N. (1998): *Basic Econometrics*. McGraw-Hill, New York, 2nd edn.

Gustaffson, B., and M. Johansson (1999): "In Search of Smoking Guns," *American Sociological Review*, 64, 585–605.

Jovanovic, B. (1998): "Vintage Capital and Inequality," *Review-of-Economic-Dynamics*, 1(2), 497–530.

Kuznets, S. (1955): "Economic Growth and Income Inequality," *American Economic Review*, 45(1), 1–28.

Leamer, E. E. (1978): *Specification Searches*. John Wiley, New York.

Levine, R., N. Loayza, and T. Beck (2000): "Financial Intermediation and Growth: Causality and Causes," *Journal of Monetary Economics*, 46(1), 31–77.

Levine, R., and D. Renalt (1992): "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82(4), 942–63.

Li, H., L. Squire, and H. Zou (1998): "Explaining International and Intertemporal Variations In Income Inequality.," *The Economic Journal*, 08(446), 26–43.

Madigan, D., and J. York (1995): "Bayesian Graphical Models for Discrete Data," *International Statistical Review*, 63, 215–232.

Mankiw, G. N., D. Romer, and D. N. Weil (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107(2), 407–37.

Raferty, A., D. Madigan, and J. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, 179–191.

Raferty, A. E. (1995): *Bayesian Model Selection in Social Research*Sociological Methodology 1995. Blackwell, Cambridge.

Rao, J., and R. Tibshirani (1997): "The out-of-bootstrap method for model averaging and selection," Discussion paper, University of Toronto.

Sala-i-Martin, X. (1997): "I Just Ran Two Million Regressions," *American Economic Review*, 87(2), 178–183.

Smith, M., and R. Kohn (1996): "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317–343.

Sokoloff, K. L., and S. L. Engerman (2000): "Institutions, Factor Endowments,and Paths of Development in the New World," *Journal of Economic Perspectives*, 14(3), 217–32.

Tierney, L., and J. B. Kadane (1986): "Accurate approximations for posterior moments and marginal densities," *Journal of the American Statistical Association*, 81, 82–86.

Western, B. (1996): "Vague Theory and Model Uncertainty in Macrosociology," *Sociological Methodology*, 26, 165–192.

World Bank (1998): "World Development Indicators CD," .

Zhu, S. C., and D. Trefler (2001): "Ginis in General Equilibrium," NBER Working Paper 8446.

Table 1. Cross-Country Regression Specifications for the Gini Coefficient

A Comparison of Regression Specifications in the Empirical Literature

	Li, Squire, and Zou(1998)		Bjorklund and Johansson (1999)		Barro (1999)		Benabou (1996)	
Covariates	coefficient	t-stat	coefficient	t-stat	coefficient	t-stat	coefficient	t-stat
GDP/capita			-0.67	0.246				
ln(GDP/capita)					41.5	0.08		
ln(GDP/capita)^2					-2.54	0.53		
avg. years primary					-1.61	0.37		
avg. years secondary	-4.55	0.78			-1.09	0.7		
avg. years higher					8.2	3.4		
civil liberties / democracy	1.61	0.3			-0.3	1.5		
land Gini	0.16	0.02						
M2 / GDP	-7.73	2.44						
inflation			-0.084	0.048				
% in industry			-0.452	0.101				
% imports from LDCs			1.34	0.487				
% unionized			-0.057	0.03				
% unemployed			-0.144	0.098				
% under age 15			0.434	0.118				
% over age 65			-0.244	0.279				
% public consumption			-0.363	0.114				
% social security transfers			0.078	0.131				
initial Gini coefficient							-0.015	0.0005
# observations	166		89		225		69	
R^2	0.62		0.69		0.56 - 0.67		0.06	
econometric technique used for estimates	OLS (IV gave similar results)		Country Fixed Effects		SURE for 4 decades		dependent variable is ΔGini/T	

The land Gini coefficient is a fixed country-specific variable. Li, et. al. (1998) also treat average years of secondary schooling as a fixed factor, using the 1960 value. The democracy variable used by Barro (1999) is the civil liberties used by Li, et. al. combined with an additional measure of political rights.

Table 2. List of Inequality Covariates

Quantity of Interest	Variable	Obs	Mean	Standard Dev.	Min	Max	Obs available for regression 8(a)	8(b)	mod	Source
(1)	(2)	(3)	(4)	(5)	(6)	(7)	8(a)	8(b)	(9)	(10)
Income Gini	adjgini	509	39.5	11.0	17.3	67.8		509		DS / WIID
Per-capita GDP	pcgdpsh	445	5057	4367	316	18935		445		World Bank
Distribution of Schooling	psyrs	413	5.1	2.5	0.1	11.2		394	*	Barro - Lee,
	hyr	417	0.2	0.2	0.0	1.5		394		UNESCO
% Under 15 / Over 65	under15	499	0.3	0.1	0.1	0.5		386		World Bank
	over65	499	0.1	0.0	0.0	0.2		386		
Agricultural Share	agshare	495	35.2	26.2	0.2	102.5		381	*	World Bank
Industrial Share	indshare	475	28.8	14.0	0.4	66.9		361	*	World Bank
Government Consum.	govc	489	14.4	5.4	3.2	38.6		354	*	World Bank
	gc	428	14.1	5.4	3.2	38.6	(334)			
Private Credit	credit	392	37.2	29.4	0.2	200.6	(308)	318	*	World Bank / LLB
Investment	gdi	447	23.2	7.0	4.1	47.1		318		World Bank
Democracy - autocracy	dma	436	2.6	7.3	-10	10		312		Polity IV
	dma_lny	411	26.3	60.8	-84.76	98.49		312		
Trade / GDP	trade	432	64.7	49.6	5.3	439.6		309		World Bank
Ag VA/employee ratio	agvapc	372	0.7	2.4	0.0	35.3		279	*	World Bank
Birthrate (crude, per 1,000)	birthrate	359	23.1	11.4	8.5	56.8		189		World Bank
Deathrate (crude, per 1,000)	deathrate	357	10.0	3.8	3.6	29.8		189		World Bank
Inflation	inflation	367	53.6	233.9	-2.98	2569.1		180		World Bank
Income Taxes (% of exp)	taxesinc	265	28.3	16.8	0.0	78.0	(149)			World Bank
	transfers	249	37.7	21.3	1.1	74.9	(145)			World Bank
	socsectax	265	12.3	14.7	0.0	54.2	(149)			World Bank
	edspending	189	4.7	2.0	0.9	10.7	(107)			World Bank
Land Gini (fixed)	landgini	280	64.1	16.0	33.85	92.3	(135)			Deininger - Olinto
	landgini_lny	279	527.5	145.0	230	826	(135)			
Terms of Trade growth	wltot	396	-0.01	0.19	-1.00	0.95		173	*	World Bank / LLB
Real Interest Rate	rint	285	0.01	0.11	-0.47	0.47		141	*	World Bank / LLB
Revolutions & Coups	rc	499	0.53	1.34	0	13		141	*	World Bank / LLB
Ethnic Fractionalization	ethfrac	376	37.6	29.5	0	93		130		World Bank
Assassinations	assass	470	1.04	3.64	0	41		130		World Bank
ln(black market premium)	lbmp	340	1.81	1.93	-2.04	9.54		125	*	World Bank / LLB
ln(population)	lnpop	507	16.3	1.6	11.1	20.9		125		World Bank
ln(population density)	lndens	468	-5.1	1.9	-10.3	-0.6		125		World Bank
ln(area)	lnarea	469	11.3	1.6	5.35	13.79	(125)			World Bank
% urban	urban	507	53.4	23.4	5	100		125		World Bank
per-capita income growth	pcgrow	411	2.0	3.9	-25.1	12.7		125		World Bank
contract enforcement index	enf	508	0.63	0.20	0.29	1		125		LLB
corruption index	crpt	508	0.57	0.24	0.10	1		125		LLB
property rights index	prop	508	0.54	0.35	0	1		125		LLB
foreign direct investment flows	fdi	345	1.11	1.77	-6.99	12.57		119		World Bank
% muslim in 1980	muslim80	285	8.5	22.6	0	96.8		107		LLB
% catholic in 1980	catho80	285	44.1	40.0	0	96.9		107		LLB
tropical dummy	tropical	508	0.39	0.49	0	1	(107)			World Bank
landlocked dummy	landlock	508	0.18	0.38	0	1	(107)			World Bank
transition (1995 only)	transit	509	0.04	0.19	0	1	(107)			World Bank

Table notes: first column lists quantity of theoretical interest, second column gives the name of the variable representing that quantity, third column the total number of observations on that variable. The fourth through seventh columns present summary statistics. The eighth column shows the total number of observations available for a regression containing that variable and all of the ones above it. Numbers in parentheses (under heading (a)) are used for illustrative purposes only, and are not included in the observation counts for the variables below it. An asterix (*) in the ninth column indicates that modifications were made to the raw data increase the number of available observations, for example, by combining information from two datasets, by using values from previous periods, or by using fitted values. The final column lists the primary source(s) of the variable.

Sources: World Bank data from Easterly and Sewadeh (2001) and/or World Bank (1998); Barro and Lee (2000) data on schooling years extended using UNESCO data on schooling duration; **?** based calculated data from FAO World Census of Agriculture; LLB=Levine, Loayza, and Beck (2000), ASB=Arthur S. Banks Cross National Time-Series Data Archive.

Table 3. Comparison of Data Sets using the Barro (1999) specification

Determinants of inequality: Barro (1999) specification, random effects

variable	Barro (1999) reported	Random Effects (RE) w/ Barro Gini data		RE w/ New Gini Data 10 years	RE w/ New Gini Data 5 years	*Adjusted Gini data* 5 years
ln(GDP)	41.5	41.12	40.12	36.87	31.09	31.81
	0.084	*10.77*	*8.50*	*9.66*	*8.22*	*8.01*
ln(GDP) squared	-2.54	-2.53	-2.45	-2.20	-1.82	-1.89
	0.53	*0.67*	*0.54*	*0.61*	*0.52*	*0.50*
Dummy: net inc or expenditure	-4.96	-3.98	-4.92	-5.75	-1.96	
	0.94	*1.03*	*0.86*	*1.04*	*0.68*	
Dummy: personal inc.	-1.19	-1.09	-0.30	0.69	1.22	
	0.87	*0.91*	*0.75*	*0.93*	*0.70*	
Avg. years primary	-1.61	-1.46	-1.44	-1.87	-1.75	-1.74
	0.37	*0.54*	*0.50*	*0.53*	*0.50*	*0.49*
Avg. years secondary	-1.09	-0.75	-1.45	-1.33	-2.00	-1.77
	0.7	*0.74*	*0.69*	*0.76*	*0.68*	*0.66*
Avg. years higher	8.2	9.96	9.45	5.96	5.36	5.64
	3.4	*3.11*	*3.07*	*3.48*	*3.08*	*3.00*
Dummy: Africa	11.3	14.43	12.28	12.50	10.83	13.03
	1.5	*2.09*	*1.97*	*2.27*	*2.06*	*2.01*
Dummy: Latin America	9.2	10.05	9.04	8.88	10.25	9.61
	1.2	*1.75*	*1.74*	*1.78*	*1.76*	*1.72*
Democracy	-0.3	-0.03	--			
	1.5	*2.69*	--			
# of observations	225	195	235	201	340	340
R - sqaured (overall)	.56 -.67	0.64	0.62	0.63	0.54	0.56
sigma u / sigma e	--	0.68	0.70	0.69	0.73	0.74

Column 1 reports results printed in Table 6 of Barro (1999). Columns 2 &3 represent our attempt to replicate these results using data constructed based on the information in Barro (1999). Columns 5 & 6 represent regressions run using unadjusted data constructed following the method outlined in Section 3.1. Column 7 represents data adjusted for survey type as described in the Appendix.

A Notes on the Construction of the Income Inequality Data Set

Cross-country comparisons of inequality are complicated by substantial variation in survey methods. In particular, income surveys differ according to whether they use (i) income, earnings, or expenditure data; (ii) household or personal income data, and (iii) gross or net (after-tax) income. Deininger and Squire (1996) outline a set of standards for considering their data to be "high-quality"—that they be household level surveys, reporting gross income, with national coverage—and these have been adopted as the benchmark in most subsequent research. Nevertheless, a non-trivial amount of variation exists in survey types remains in their dataset, and typically individual researchers have taken different approaches to correcting for these differences. Bénabou (1996), for instance, uses dummy variables in his regressions to control for survey type, while Forbes (2000) uses a simple fixed correction, adding 6.6 to Ginis constructed using expenditure data.

The dataset used in this paper was constructed by a very rigorous process of screening and correcting observations to ensure the greatest comparability. Improving ex-ante comparability among the observations was eased as observations were drawn from a more comprehensive pool than that used in earlier studies. Whenever possible, the maximum number of observations were drawn from the same source.[36]

Both as a benchmark, and to improve upon the rather crude "usual corrections"country-specific fixed-effects regressions were run on the full **WIID** database (nearly 3,000 observations) to estimate the effects of each survey type. The **WIID** provides a more detailed breakdown of survey methods than does **DS**). This approach allows for ex-ante "corrections" for survey type similar to that used by Forbes, but with a higher degree of accuracy – both by allowing for additional hedonic characteristics of surveys beyond just expenditure, and by estimating the effect of surveys using only variation within each country. Estimates based on cross-sectional variation are likely to biased, since survey methods appear to be correlated with levels of inequality[37] The corrected Gini, the variable "*adjgini,*" is used in all the following tables and analysis.

A.1 How data set was constructed

The three priorities guiding construction of the income inequality data set were comprehensiveness, consistency and accuracy. Choices made regarding which observations to use were made by weighing these priorities. To meet the first priority of comprehensiveness, all available observations that were deemed "OK" by the WIID were considered. The **DS** "high

[36] This generally applied to the time series within countries, but in some cases also applied to authors who conducted surveys in multiple countries.

[37] This is not a theoretical but a practical point: survey types are often correlated with particular regions. For example, surveys measuring net personal expenditure happen to be prevalent in (relatively high-inequality) African countries.

quality" data set, which is a subset of the WIID database, was taken as a starting point, and additional observations were added from the WIID when there were none present in the **DS** database.

To meet the second priority of consistency – first within countries across periods and secondly across countries – the priorities of Deininger and Squire were followed, and, where possible (in most cases) their observations classed as "acceptable" (or "acc") were used. In almost all other cases, it appeared that the multiple candidate Ginis from the WIID were sufficiently similar that empirical results would not likely be sensitive to alternative choices.

Although more complete than the datasets built on **DS** by Forbes or Barro, the inequality panel still contains numerous gaps in years covered for many countries. Because inequality data often are available for countries only for different years, some grouping of observations across a range of years is necessary to create a sufficiently large cross-section. Given that the focus of attention are longer term trend movements, this is not too troubling. However, it requires taking a stand on the question of the appropriate time span necessary for capturing basic trends of interest. Different approaches to this dilemma have been taken in the literature previously. Bénabou (1996) averaged all available observations within a decade (in part to reduce measurement error), while Forbes (2000) designed a panel based on 5 year intervals, using observations from previous years when one for the target year was not available. Barro (1999) employed but a similar approach, but suggested 10 year periods are more appropriate for considering changes in inequality. Without taking sides on this issue, but assuming more data is better than less, the dataset used in this paper is based on a panel of 5 year increments. When data for a particular year was not available, the closest neighboring observation was used. If multiple neighboring observations were available, the target year was interpolated.

Taking the **DS** criteria as an accepted benchmark and their dataset as a starting point, I constructed the dataset used in this paper from the following steps:

1. All "OK" observations were drawn from the WIDER database (2,976 obs. total). These include 4 types of observations not given an "acc" (high quality) rating by DS:[38]

i) non-national surveys (rural, urban, metropolitan, taxpayers only, etc.) ("nn" in DS set)

ii) observations superseded by the existence of other data based on consistent surveys ("cs" in DS set)

iii) observations excluded for lack of a primary source ("ps" in DS set)

iii) observations not included at all in the DS dataset at all.

2. Data for each country was reviewed individually, to establish what surveys were used and the comparability of their methods and the resulting Gini measurements. Based on

[38] The shortcoming of non-national surveys is obvious, and observations rated "nn" were used only in a few occasions when the data were deemed preferable to including no information (e.g., Argentina, which only surveys in cities). Given a fixed-effects specification and panel estimation techniques, consistency of surveys across countries was deemed more important for the purpose of tracking convergence than complete comparability across countries. Problems with the latter three classifications are less immediately obvious, however, and observations of these types were included on occasion.

this information, the most reliable and most comparable series were used to create up to 8 observations (1960, 65, ..., 95) for each country. No adjustments were made for differences in survey types (income/expenditure, gross/net, household/individual, etc.) although that information was included for later, data-based adjustments.

3. The resulting datasetwas compared with the **DS** database and differences in choice of Gini were explored and reconciled. **DS** "acc" observations were given second priority, after within-country compatibility of time series observations. The final variable used either the **DS** observation or the WIID database. The latter observations were notated by using the variable "W" in the field `wdtype`.

In the general trade-off between measurement error (from lower quality data) and greater cross-country variation, priority was given to the latter. The goal was to maximize the number of observations through the inclusion of additional (generally less developed) countries and in the number of years covered. Unlike previous work, which has treated inequality as an independent variable partially determining growth rates, the present research is focused on inequality as the dependent variable. This means that measurement error in the Gini coefficient is much less troubling (as the cost is only lower standard errors, not biased coefficients).

A.2 Gini Adjustments

Although the dataset was generated to ensure the closest compatibility of observations across time, there are significant differences in survey methodology across countries that make cross-country comparisons of Gini coefficients troublesome. The benchmark Gini coefficient is measured using gross household income. To aid in adjusting Gini coefficients calculated using surveys that depart from this methodology, three dummy variables were created

1. `expdum` survey measured expenditure, not income
2. `netdum` survey used net, not gross income (expenditure)
3. `perdum` survey studied persons (or equivalents) rather than households/families

A priori, one might expect that income Ginis measured from expenditure would be lower than for income, due to consumption smoothing behavior. Similarly, one would expect that progressivity inherent in most government policies would lead net incomes (or expenditure) to be flatter over the population than gross incomes (and generate a negative correlation). The correlation between measured inequality and the unit of observation (person or household) is less clear. If those with higher incomes enjoy larger families, per-capita (or any other equivalence scale) incomes will tend to converge. On the other hand, the potential for assortive mating could generate more inequality in household incomes than personal incomes.

The following table captures the simple correlations between the three basic survey types and measured income gini coefficients. It shows that lower measured inequality is associated with net and personal measurements, but almost no correlation exists between the use of expenditure or income and measured inequality. This does not justify its exclusion as an important variable, however. The correlations between dummy variables reveals that surveys measuring expenditure often also happen to use net and personal measurements. This highlights the importance of using a multivariate approach to control for each effect individually.

($obs = 393$)	`hopgini`	`expdum`	`netdum`	`perdum`
`hopgini`	1.00			
`expdum`	0.03	1.00		
`netdum`	−0.35	0.48	1.00	
`perdum`	−0.12	0.39	0.29	1.00

Four approaches were attempted to correct for measurement error introduced by survey type. The first was to run a simple regression on the dummy variables using the 393 observations in the dataset calculated in this paper, compute adjustments and generate an "adjusted Gini" variable. The second was to do the same thing using the largest possible number of observations, in this case the WIID dataset with 2,966 observations, to calculate the corrections. The third was to run these regressions using a fixed-effects panel data estimator. This technique should much more accurately measure the survey effects because much of the variation in survey techniques is across countries, and estimation in a cross-section may be biased if the survey method is correlated with the level of inequality. The panel data method properly attributes most cross-country variation to individual country effects, and estimation of differences in surveys is captured through differences over time within countries. The consistency of inequality over time provides some assurance that measured variance within countries is due to differences in survey types, and not actual changes in inequality. The final approach used was to use unadjusted Ginis in the regressions, and to include the dummy variables in the regression equations themselves.

The following results summarize the measured survey effects from different regressions. Columns (1) and (2) are pooled and fixed effect (FE) estimates using the 393 "high quality" gini coefficients used in this paper. Column (3) additionally controls for the education distribution. The final column (4) is analagous to table (2) but using the full WIID gini database.

Survey Control	(1)	(2)	(3)	(4)
`expdum`	7.49	−4.05	3.01	−5.08
`netdum`	−9.94	−1.62	−7.84	−1.96
`perdum`	−1.97	0.39	−3.38	4.66*
[education variables included]	no	no	yes	no
no. obs.	393	393	350	2,976
regression type	pooled	FE	pooled	FE

*the WIID database has a much more detailed breakdown of survey types than the DS database, and regressions were run using all controls separately. The reported coefficient on perdum is actually the sum of the coefficients on sample (4.76),enum (.58) and unit (−.68). In general, these tend to all be 1 or 0 together, but the logical relationship between each is (sample$=1 \Longrightarrow$ enum$=1 \Longrightarrow unit = 1$). Additional controls for related variables were enum (which is 1 if the survey collected personal, not household data) in the WIID (sample$=1$ implies enum$=1$, but not vice versa).

The variation in estimates calls into question the robustness of any one set of results.

Methodologically, the fixed-effects estimator would seem to guarantee the most accurate results, and use of the full WIID database–even if it includes some lower-quality data–provides a much larger sample of variation of survey types within countries and therefore much smaller standard errors. In making adjustments, there is an assumption that survey effects are the same across countries–an assumption that may or may not be innocuous. If survey effects do vary across countries, using estimates from columns (1)-(3) are preferable over column (4) which draws estimates from a large set of Gini coefficients not used in the rest of the analysis. However, the similarity in results in columns (2) and (4) offers some support for the robustness of the coefficients at least for `expdum` and `netdum`. The large variation in perdum makes any significant adjustments more questionable.

Some jugdement over these adjustments is necessary, and this paper adopts two approaches. In linear regressions, the favored technique was to use the unadjusted Gini coefficients and then to control for survey effects within each regression. When these adjustments could not be made, the following adjustments were used to create the variable "AdjGini":

if `expdum`= 1 add 4.5 to hopgini (average of columns (2) and (4))
if `netdum`= 1 add 1.75 to hopgini (average of columns (2) and (4))
if `perdum`= 1 subtract 0.5 from hopgini (weighting column (2) results relative to (4))

By way of comparison with other work, Forbes (2000) makes only a single adjustment, adding 6.6 to Ginis measured using expenditure. In my sample, 83% of those cases (65 of 78) the Gini is also measured net, and for 69% both `netdum` and `perdum` equal one. The table below gives the percentage of the 78 observations with different adjustments. In short, the adjustments made are similar to those of Forbes and other authors, but with slightly more flexibility.

if exp=1 and...	the adjustment is	in _ _ of the 78 cases
net=0, per=0	+4.5	5%
net=0, per=1	+4.0	9%
net=1, per=0	+6.25	14%
net=1, per=1	+5.75	69%

B Some considerations for the cross-country empirics of inequality

This section presents a quick overview of issues surrounding these decisions. Issues of variable selection and model specification will then occupy the remainder of the paper.

B.1 Time-series versus Cross-sectional Inequality

Although each of the authors cited above utilize longitudinal data in estimating the impact of inequality, Bénabou and GJ study the average time series behavior of inequality across a set of countries, while Barro and LSZ focus principally on the determinants of cross-sectional variation. Asking why inequality measures move in a time series and why they vary in a cross-section of countries are very different questions, and as our discussion of theory suggests, the answers may or may not similar.

In theory, one can imagine at least four reasons why we might see inequality vary more in a cross-section than a time series. The first is that inequality depends on some vector of covariates X_{it} and these vary more in a cross-section than in a time series. Alternatively, inequality may depend on both X_{it} and some *country-specific* characteristics Z_i, or these country specific characteristics Z_i modify the marginal impact of X_{it} on the Gini coefficient. Finally, it may be that the available data is insufficient to answer the question: there is some country or region-specific effect μ_i that can be identified, but not explained by our set of covariates.

These hypotheses can be nested in the following regression equation

$$G_{it} = A'X_{it} + B'Z_i + (X_{it})' CZ_i + \mu_i + \eta_{it} \tag{5}$$

where η_{it} represents a conventional white-noise term. Of course the dimensionality of the covariate vector relative to the number of available observations is potentially quite large, particularly when including the matrix C interaction terms or allowing for higher order powers to capture potential non-linearities. This presents a challenge to estimating equation (5) and then attempting formal hypothesis testing. The size of the data sample that can be employed in cross-country regressions is also dependent on the number of included observations (Section 3 and Table 2). Therefore, some combination of *a priori* reasoning and iterative analysis of the data is required in the search for a correct specification. A discussion of the methodology is presented in Section 4.

B.2 Econometric Techniques for Panel Data

The simplest method of estimating the coefficients A, B, and C in equation (5) is simply to pool observations across countries and over time and then perform ordinary least-squares. Of course, disregarding the identity of country and time period seems to entail the loss of potentially important information. Whether or not the country and time period *do* provide any useful information depends on the *exchangeability* of the errors, which in turn depends on whether an appropriate conditioning structure has been adopted.[39] This, of course, is one of the central issues in model choice.

The easiest way to ensure that all country-specific information is incorporated into the conditioning structure is through the introduction of country specific effects.[40] These can be fixed effects, using dummy variables (the coefficients of represent estimates of μ_i), or random

[39] If country identity contains information for predicting inequality even after conditioning on all observable coefficients, then the assumption of conditional exchangeability is violated (see Brock and Durlauf (2001) for a discussion). Exchangeability is fundamental to the probablistic characterization of the data as independent draws from a statistical distribution.

[40] Time-specific effects can be introduced similarly. However, because there does not seem to be any sys-

effects, in which the unobserved country-effects are treated as draws from a distribution with estimated variance σ_μ^2. The advantage of fixed effects is that no assumptions regarding the independence between the covariates and the unobserved μ_i are necessary, so it is a potential solution to the problem of omitted variable bias. The cost, however, is that *all* country specific variables must be omitted, making it difficult to assess the impact of individual elements of Z.[41]

Averaging over t and subtracting from equation (5) gives the country fixed-effects (or "within") specification

$$\left(G_{it} - \bar{G}_i\right) = A\left(X_{it} - \bar{X}_i\right) + C\left(X_{it} - \bar{X}_i\right) Z_i + \left(\eta_{it} - \bar{\eta}_i\right) \tag{7}$$

where a bar over a variable represents the average over time within each country.[42] Note that the direct impact of fixed covariates Z also drops out in the differencing operation, so although the coefficients in C can be estimated, those in B cannot. From the standpoint of efficiency, given the relatively short time dimension of the panel for most countries, introducing fixed country dummies also significantly reduces the available degrees of freedom.[43]

Despite their apparent flexibility in dealing with unknown sources of heterogeneity, it should be noted that both of these estimators presume the marginal effect of covariates over time is the same as that across countries. Consider the role of log(GDP) in a fixed effects regression, for example: the within estimator averages the response of changes in inequality to changes in log income, or economic growth. Estimating the same relationship in a pure cross-section typically generates very different estimates (e.g. Barro (1999, Table 4)). In this case, log(GDP) represents the relative position of the country within the international distribution of income – which has a different set of theoretical implications for inequality (regarding, e.g. the effects of trade theory, distance from the technological frontier, and the

tematic additive shocks to the distribution of inequality from period to period, the discussion of groupwise heterogeneity is focused on country-effects.

[41] A compromise between a pure cross-sectional regression of country averages and the fixed effects approach is to use a random effects specification, which estimates coefficients using a weighted average between equations (5) and (7):

$$\left(G_{it} - \alpha\bar{G}_i\right) = \left(A + CZ_i\right)\left(X_{it} - \alpha\bar{X}_i\right) + (1-\alpha)\left(BZ_i + \mu_i\right) + \left(\eta_{it} - \alpha\bar{\eta}_i\right) \tag{6}$$

The weight α reflects the ratio of the variances of μ_i and η_{it}, which are parameters to be estimated. Although this approach allows estimation using both country-specific knowns Z_i and unknowns μ_i, it is clear from equation (6) that identification of the coefficients requires the random effects μ_i be uncorrelated with the covariates. This is unlikely to be the case in cross-country regressions, however.

[42] Least-squares estimation of (7) is equivalent to allowing a dummy variable (or fixed effect) for each country to estimate μ. Although the country specific terms do not appear in (7), the dummy coefficients μ_i are easily obtained once $\hat{A}$ is estimated.

[43] The number of Gini coefficient observations in the dataset averages 3.8 per country (although some countries have as little as one). Using fixed effects therefore entails estimation over only about 2.5 observations per country at best. In practice, due to limited covariate data, even fewer available observations will result.

like). A more complete model might therefore specify

$$G_{it} = \beta_0 + \beta_1 \left[\log(GDP_{it}) - \log(\overline{GDP}_i)\right] + \beta_2 \log(\overline{GDP}_i) + \varepsilon_{it}$$

where β_1 represents "short run" effects of income growth within a country, and β_2 captures the "long run" effects of changing a countries' relative position in the world's (per-capita) income distribution. For most variables, introducing this level of flexibility is not likely to merit the additional parameters required, so use is limited to those cases (such as per-capita income) with some reasonable theoretical justification.

B.3 Data Choice: Quality versus Quantity

The bulk of recent research on inequality has employed the **DS** data set, which covers a wide range of countries at the cost of introducing non-trivial measurement error from the variation in survey methodologies. Notably, however, GJ instead use data from the Luxembourg Income Study, which has harmonized microeconomic data sets for 16 OECD countries. Despite the reduced cross-sectional variation, the authors are thus able to utilize data less subject to measurement error, which is tailored to their specific interests ("person equivalent" disposable income), in a more complete panel. This trade-off of quality of data for quantity of coverage seems appropriate given the authors' focus on within-country changes in inequality over time, which are relatively small. In addition to maximizing the signal-to-noise ratio, the likelihood that the estimated marginal impact of covariates differs significantly across samples is much lower.

At the other end of the methodological spectrum is Barro (1999), who expands upon the set of traditionally used "acceptable" **DS** Gini coefficients to include additional observations from developing countries that were previously excluded due to lack of a primary source. Barro argues that the additional cross-sectional variation from these data points justifies the potential cost in terms of increased measurement error.

Attempts to evaluate among competing models are further complicated by the missing data for many countries and time periods. In general, the number of observations available for regression (the intersection of observations with non-missing data on all regressors) falls quickly as the number of regressors is increased (Table 2). This complicates conventional approaches to model selection which involve "nesting and testing" specific subsets of coefficients in a large model including all potential regressors, since even smaller models can be severely constrained by a few variables in terms of observations available. The implication is that practical considerations dictate variable choice as much as theoretical considerations. We will return to discussion of this problem later in the context of model selection. The first challenge, however, is to generate a data set that is as complete as possible, containing observations of the highest comparability between and within countries.

C Practical Challenges to Specification Search: Details

C.1 Data imputations

Distribution of Economic activity. Employment shares are a natural interpretation of the theoretical quantity λ described in Section 2, the distribution of activity measured by value added (VA) as a share of GDP also incorporates information on productivity (i.e., underemployment in agriculture). As a result, the variables `agemploy` and `agvapc`=agricutural VA/`agemploy` as separate variables. Industrial VA was also employed. To deal with missing obs, when data on `agemploy` was unavailable, fitted values (based on agricultural VA and other controls) were used. A similar apprpoach was used for industrial VA based on fitted values of industrial employment share for males.

Real interest rate was measured using a 5-year prior period average of data from the World Bank 1998 World Development Indicators (WDI). When missing, observations on the variable `brra` were taken from Levine, Loayza, and Beck (2000) (LLB). In five instances, the real interest rate was calculated from the lending rate minus inflation.[44]

Private credit. Data were taken from the WDI, with missing observations replaced with `privo` from LLB, the period average of private credit, measured using IFS data. In the case of Romania, 1990, 1995, data on bank credit was used.

Institutional variables were taken from LLB, an average of 1982-1995 (the widest span of years available). These include enforcement (`enf`), a measure of government's ability to abide by rule of law (an average of LLB's measures of rule of law and measure of contract enforcement), corruption (`crpt`), and an index of property rights (`prop`). All of the original indexes were transformed to a scale of 0 to 1. Missing observations were imputed using controls for region, income level, and legal origin, log area, and whether the country was in the tropics. One difficulty is that no observations were available for socialist countries. In the case of property rights, all socialist countries were assigned a value of 0.

C.2 Proxy searches

Kuznets/structural adjustment can be measured using any two of three sectoral shares (agriculture, services, and industry). Regressing the Gini on these shares returns coefficients of roughly .49 for services and agriculture, and .29 for industry. This implies that for every 1% of activity that shifts into industry, the Gini coefficient falls by roughly 20 points. Obviously employment is a more appropriate unit for assessment of income distribution than value added. Functionally, entering agricultural employment with a quadratic term gives clear Kuznets curve. Employment shares in agriculture were preferred to value added based both on theory, and because for agriculture, using employment places fewer restrictions on available observations.

[44] The new observations for the real interest rate were for Germany 1980,85,90; Japan 1960; Mex 1980; Uganda 1980.

For **financial links**: nominal interest rates, inflation, and the real interest rate are all clearly linked. Settled on using real interest rate as measure of the long-run cost of capital, and inflation rates (the average of the 5-year period). Private credit was preferredd to measures of bank credit based on theory.

Government policies: in assessing the government's role in public finance, chose between variables on income taxes, social security taxes, and transfers – using income taxes as broad measure of theoretical quantity of interest with a reasonable number of observations. Although transfers had fewer observations available, it was also used because it most closely matches the variable used in the specification of Gustaffson and Johansson (1999).

Events: In looking at the role of political events and uncertainty on income inequality, many different measures were available, including assassinations, revolutions, coups, cabinet and/or constitutional changes. The approach taken in the paper was to employ a single measure of political instability – revolutions & coups – which seemed to require sufficient predetermination that they reasonably would not be triggered by a sudden rise in inequality (and therefore be endogenous in the within-country estimation). Assassinations was also used as a measure of the level of violence and rule of law, although it was not found to be particularly signficant across a range of specifications.

C.3 Correlation vs. Causation

Inference from statistical models assumes a certain direction of causality from independent to dependent variables, both with respect to restrictions on the correlation of the error with the covariates required for statistical identification, and with respect to inference regarding policy. Uncertainty over causality is always a concern in cross-country regressions, however, requiring any such analysis to rely on theoretical reasoning and a tolerance for causal ambiguity in drawing inference. From a statistical standpoint, we need be concerned only when shocks to the level of inequality could be correlated with right hand side variables.

One example of where it might be a concern in our present analysis is in the role of political variables. In particular, although we hypothesize that inequality may worsen in situations of political and social instability, sudden rises in inequality also have the potential for triggering civil unrest. The role of the "events" variables on inequality were therfore estimated using both standard fixed effects and instrumental variables fixed effects estimation. As shown in Figure 9, employing the latter dramatically reduces the significance of the reported coefficients. However, this could reflect either a correction in existing bias, or simply due to the effect of the loss of precision on the coefficient standard errors.

The regressions reported contained no other covariates, to maximize the sample size. The same procedure was used with the full sample, however, with similar results although in this case the initial t-statistics were not significant to begin with, and instrumenting caused an initially negative coefficient on REVCOUP to rise slightly rather than fall.

Controlling for the endogeneity of political events

	standard fixed effects	instrument sample only	instrumental fixed effects
assass	-0.16	-0.11	0.73
t-stat	-2.10	-1.82	0.20
revcoup	0.32	0.04	-6.18
t-stat	1.48	0.18	-0.35
govtcrises	0.30	0.22	-1.35
t-stat	1.50	0.81	-0.22
(constant)	39.22	41.47	45.06
t-stat	136.56	131.86	5.25
obs	464	203	203

instruments include: radios, life expectancy, price of food

Figure 9: