SIEPR Discussion Paper No. 07-52

The Political Economy of Heterogeneous Development: Quantile Effects of Income and Education

By
Marcus Alexander
Harvard University

Matthew Harding
Stanford University

Carlos Lamarche
University of Oklahoma

August 2008

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Political Economy of Heterogeneous Development: Quantile Effects of Income and Education*

Marcus Alexander†, Matthew Harding‡ and Carlos Lamarche§

August 2008

Abstract

Does development lead to the establishment of more democratic institutions? The key to the puzzle, we argue, is the previously unrecognized fact that based on quantitative regime scores, countries over the past 50 years have clustered into two separate, very distinct, yet equally-common stages of political development — authoritarian states with low levels of freedom on one side and democracies with liberal institutions on the other side of a bimodal distribution of political regimes. We develop a new empirical strategy — exploiting exogenous world economic factors and introducing new panel data estimators — that allows for the first time to estimate the effects of development as well as changing unobserved country effects in driving democracy at these different stages of political development. We find that income and education have the least effect on democracy when authoritarian regimes are consolidated and that only changing country effects, possibly accounting for institutional legacies, can lead to political development. Ironically, it is in highly democratic and wealthiest of nations that income and education start to play a role; however greater wealth and better educated citizenry can both help and hurt democracy depending again on what the country's institutional legacies are. Far from accepting the notion that much of the developing world is cursed by unchanging and poor long-run institutions, policy-makers should take note that with democratization we also see changing country-specific factors that in turn condition the difference income and education make for democracy.

JEL: C13, C23, P16, O10
Keywords: democracy, economic development, quantile regression

*We are grateful to Tim Hyde for excellent research assistance. We are grateful to Robert Bates, Victor Chernozhukov, Jorge Dominguez, Dawn Brancati, Margarita Estavez-Abe, Kevin Grier, Yoshiko Herrera, Nahomi Ichino, Torben Iversen, Gary King, Roderick MacFarquhar, Thomas Remington, James Robinson and seminar participants at Harvard and Stanford for useful comments, and to Daron Acemoglu, James Robinson, Simon Johnson and Pierre Yared for sharing their historical data.

†Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138; Email: malexand@fas.harvard.edu

‡Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: mch@stanford.edu

§Department of Economics, University of Oklahoma, 729 Elm Avenue, Norman, OK 73019; Phone: (405) 325-5857; Email: lamarche@ou.edu

1. Introduction

Western history is often presented as the struggle of freedom and democracy against the forces of authoritarian government, going back as far as Herodotus who used this argument to motivate the clash between the Hellenes and the Persians. Today democracy is also associated with economic prosperity as political and economic development appear to be positively correlated over long periods of time. The failures of modern totalitarian regimes in Germany, Russia and China had fueled hope for proponents of democracy, as the last 50 years have seen stunning changes in the political landscape, from the European colonies achieving independence to the fall of the Berlin Wall. Yet, a recent report by the Freedom House, a US based non-governmental organization (NGO), documents the reversal of the trend towards democracy in 38 countries, particularly in Asia and Africa. Moreover, only two countries, Thailand and Togo have made any progress towards democracy in recent years. We show in this paper that far from being defined by the growing dominance of democracy, the second part of the 20th century was characterized by the entrenchment of two very distinct, yet equally common political regime types — liberal democracy guaranteeing political rights on one hand and autocracies characterized by poor institutions and lack of political and civil rights on the other. Analyzing the two most well-known and reliable indicators of regime type for all countries in the world since the Second World War, we find that the distribution of political regimes in the world follows a bimodal distribution.

The bimodal distribution of regime types in the world has not been previously empirically documented or explained. In addition to documenting the existence of regime clustering into two opposite types, here we also show that understanding the causes of the bimodal regime distribution holds the key to answering the question of whether development can bring about democracy. This question is at the top of international policy-making agenda, from the World Bank trying to improve institutions that maximize transparency and accountability to Western governments' foreign policies emphasizing support for democracy as both a human rights and an international security imperative. The success of both World Bank and national governments in promoting democracy in large part rests on discovering whether promoting development can bring about democracy; here, we will show that this question can be understood by analyzing and explaining why is that all countries in the world have clustered into two distinct, opposite regime types over the past half century.

While economists have emphasized the importance of democracy stimulating economic growth (Persson and Tabellini, 1997, 2000, 2003; Persson et. al. 2000), we must also address the question of whether economic development can stimulate political development. The question of how income affects democracy has only recently captured the attention of economists. In this paper, we focus on two potential forces of development, income and education, and assess the extent to which they can facilitate the struggle for civil and political liberties, as well as the establishment of liberal democratic institutions that preserve those liberties in the long run. This is relatively unexplored territory for economists, who first directed their attention to this problem after Barro (1999) showed that a number of development indicators such as improvements in living

standards, especially GDP and education, reduced dependence on natural resources, middle-class share of income and certain religions affiliations are more conducive to democratization. Many of the issues discussed in this study have been mentioned in some form or another in the vast political science literature of the past 50 years, but were approached from a less quantitative perspective. The earliest systematic formulation of the connection between political and economic development is due to Lipset (1959, 1963).

To explain the bimodal regime distribution and the underlying effect that development has on this distribution, we had to develop a new empirical strategy that simultaneously solves the problems of (1) reverse causality given that it is well established that institutions affect development; (2) unobserved heterogeneity and long-run country-specific effects given that no set of indicators in any comparative study can account for all country differences in institutions, culture and historical legacies that could affect democracy; and (3) the fact that all existing empirical strategies in the literature can only be used to estimate the effect of income on the mean of democracy, which may be uninformative given that we demonstrate that regimes are distributed bimodally. Our empirical strategy involves using three different sets of instruments that identify exogenous variation in income but are presumably not correlated with democracy. Two of these instruments (geography and trade) have been used in the literature previously, while our third and preferred set of instruments is based on a new strategy of using shocks to world markets measured by estimating global factors. We solve the second and third challenges above by designing several new estimators that can enable us to measure the effect of development at different quantiles of democracy distribution, while at the same time allowing us to identify changing country-specific factors that stand in for long-run institutions, culture and historical legacies. Therefore, the entirely new empirical approach is to estimate IV quantile effects as well as unobserved heterogeneity both at country-specific and distribution-specific levels.

The starting point for our study therefore is the finding that the distributions of two commonly used numerical measures of democracy are bimodal with most countries concentrated at the extremes. This motivates our methodology as a departure from traditional mean regression techniques and instead re-orient the topic of the discussion on the estimation of different effects of the variables of interest at different quantiles of the democracy distribution. Adding our three sets of instruments, we identify the effect of economic development that is stronger in the middle range of the distribution and almost non-existent in the tails. In practical terms we find that economic development has a three to four times larger impact in Latin America than in Sub-Saharan Africa. This finding lends support to an emerging focus on so-called "hybrid democracies," regimes that have spent years or even decades in-between the period of a collapsed dictatorship and a full-fledged democracy, including many African, Asian and Eastern European countries after the fall of communism. Being acutely aware of the difficulties typically encountered by applied researchers in finding valid instruments, we show that similar results hold when we use a set of geographic instruments, an instrument based on world trade, and a set of instruments designed to capture world economic factors. These inverted U-shaped relationships between income and democracy and between schooling and democracy

over the quantiles of the distribution of democracy are found to be robust to a number of econometric specifications.

However, our most significant finding arises only when we estimate changing country-specific effects. For the first time in the literature, we allow for the estimation of different effects that are not only country-specific but that furthermore are allowed to vary across the distribution of democracy. We find that once we account for country-specific effects, the inverted U-shaped relationship described above disappears. In fact, for the low quantiles of the democracy distribution, the effect of income and schooling on democracy is very close to zero. Hence, the answer to the question of whether development can promote the establishment of democracy in autocratic regimes is no.

Having found that country specific effects matter disproportionately more than economic development in determining democracy, we ask whether this is uniformly so over the distribution of democracy. The analysis suggests, however, that the importance of these factors diminishes as countries become more democratic. Even for democratic countries, however, we find evidence of economic development playing a heterogeneous role, a fact consistent with a large literature emphasizing institutional differences in modern democracies (Lijphart, 1999; Hall and Soskice, 2001). Institutional differences are well known to explain different cross-country differences in economic performance, particularly in the case of labor market outcomes (Blanchard, 2004). More recently differences in labor market institutions have been explained by differences in civic virtues across countries that jointly determine both political and economic outcomes (Algan and Cahuc, 2008). In this paper, we find that the country specific effects measured by the econometric model are a combination of long-run institutional factors and contemporary circumstances such as military coups or constitutional reforms.

Our paper complements the recent quantitative work of Acemoglu et. al. (2008), which analyzes similar data within the context of a mean regression framework and our findings are broadly consistent with theirs. Our quantile regression framework however allows us to explore the same issues from the perspective of increasingly fine granularity and thus direct the discussion to the role of cross-country heterogeneity and the extent to which statistical relationships are conditioned by the relative position occupied by a country in the distribution of democracy.

More than providing finer detail, as we address in our discussion, our findings call for reorienting the current debate on income and democracy. Democracy in our analysis emerges as what has been previously termed a "foundational good", or what is even better understood as a "foundational institution." We have in fact found that all country-specific, unmeasured effects can change as countries move towards more democratic regimes. Development has zero effect, but as democracy is established, our evidence suggest that other long-run institutions, historical legacies and culture change. Far from being fixed and inescapably path dependent, these country effects can change and open the door for economic development to further affect political regimes. As countries become democratic, the new set of institutions conditions the role that

wealth and education can have on democracy. Again, far from having a simple effect, greater wealth can both promote and retard democracy.

The structure of the paper is as follows. In Section 2, we introduce the data and describe the motivating puzzle of the bimodal distribution of democracy. Section 3 discusses the basic quantile regression model and characterizes the estimated inverted U-shaped relationship between income and democracy over the quantiles of the income distribution. Section 4 extends the analysis to an instrumental variables framework. Section 5 introduces the two types of unobserved effects and makes the distinction between location and distributional shifts on the quantities of interest. Section 6 presents evidence on patterns of democracy and suggests an explanation for the puzzle presented in Section 2. Section 7 concludes. Appendices A and B provide details on the computational aspects of the estimation methods, and describes the data. The online appendix available on the Web provides additional robustness checks not reported in this version of the paper.

2. **The Puzzle of the Bimodal Distribution**

We employ two measures of democracy for which data is publicly available, the Polity and Freedom House scores. The Polity (version IV) score, compiled by an academic panel at George Mason University's Center for International Development and Conflict Management, is defined as the difference between an index of autocracy and an index of democracy for each country. Each government is assigned a number between 0 and 10 on each scale based on a set of weighted indicators designed to capture the extent of competitive political participation, institutionalized constraints on executive power and guarantees of civil liberties and political participation. The primary focus of the index is on central government and it notably ignores the extent to which control over economic resources is shared and the interaction between central government and separatist or revolutionary groups. We use a sample of countries from 1945 to 1999 normalized to range between 0 and 1.

The Freedom House democracy score is compiled by a New York based NGO founded by Eleanor Roosevelt and aims to measure the extent of freedom as experienced by individuals. It consists of a rating system involving 10 political rights and 15 civil liberties questions. The questions cover diverse categories such as the electoral process, political pluralism and participation, functioning of government, freedom of expression and belief, associational rights, rule of law and personal autonomy. By design it places greater emphasis on experienced freedom as opposed to legal guarantees. We employ a sample normalized to a range between 0 and 1. The sample covers the period 1972 to 1999.

In Table 1 and Table 2, in the online appendix, we list the countries in the Polity sample with their corresponding country codes and sampling periods. The number of observations per country varies substantially with the inclusion of relatively new countries such as those previously part of the Soviet Union. In Table 3, in the online appendix, we present the summary statistics for the measures of democracy employed in this paper disaggregated by geographic regions. In spite of the different concepts of democracy that the two measures are meant to capture, from a purely numerical point of view they show remarkable agreement. As one would intuitively expect, the mean values for the Western world are about 0.9 while those for Sub-saharan Africa are only about 0.3. While these two measures rely on a substantial amount of subjectivity, they have been used extensively in quantitative studies as measures of political freedom (see, e.g., Acemoglu et. al., 2008; Barro, 1999; Fearon and Laitin, 2003).

On closer consideration of the two measures of democracy, perhaps the most striking feature is their pronounced bimodal distribution as illustrated in Figure 1 for the Polity measure and Figure 2 for the Freedom House measure. Both distributions have a mode at 1 and another mode at 0.1 and 0.2 respectively. The near lack of mass at the mean of the distribution invites the question as to the usefulness of mean regression as a quantitative tool for analyzing the relationship between income and democracy. While we acknowledge that it is difficult to make inferences based only on the unconditional distribution, we cannot but enquire whether the relationship between income and democracy differs at different quantiles of democracy. Thus, we shall depart from previous quantitative work on this subject and focus exclusively on results based on a quantile regression methodology.

Until now, few attempts have been made to explain the puzzle behind the bimodal concentration of political regimes. Recent work in political science however seems to be consistent with the emphasis we place on explaining the relationship between income and democracy at different quantiles of democracy. Przeworksi and Limongi (1997) and Przeworski et. al. (2000) argue that countries often become democratic for reasons which do not appear to be connected to income, but once they are democratic more prosperous countries are more likely to remain democratic. Epstein et. al. (2006) highlight the importance of "partial democracies", countries with fragile democratic status which tend to be volatile and highly heterogeneous. These studies appear to point out varying mechanism for political development as well a heterogeneous relationship between economic and political development at different stages of democratization.

3. **Measuring Quantile Effects**

In a typical least-squares regression model approach to the analysis of the relationship between income and democracy, we focus on estimating the best linear predictor of the conditional expectation of the dependent

variable,

$$E(D_{i,t}|I_{i,t-1}, x_{i,t}) = \gamma I_{i,t-1} + x'_{i,t}\beta, \tag{3.1}$$

where $D_{i,t}$ is the normalized democracy for country i at time t. The income variable $I_{i,t-1}$ is measured at time $t-1$. It corresponds to the log of per capita GDP. The parameter γ captures the (marginal) effect of income on democracy at the mean level. Any additional controls or variables of interest such as education or population are included in the variables $x_{i,t}$.

As we discussed in the previous section, the unconditional distribution of income is strongly bimodal with most countries clustered at the ends of the scale. Thus it seems that an analysis focused on the (conditional) mean of the distribution might miss important distributional effects of income and that by looking at the tails of the distribution we may uncover richer evidence. From an econometric point of view, a quantile regression approach may also be more robust to distributional assumptions on the error term.

3.1. **Pooled Quantile Regression Model**

We shall direct our attention to the modeling of the τ-th conditional quantile functions of democracy for country i at time t,

$$Q_{D_{i,t}}(\tau|I_{i,t-1}, x_{i,t}) = \gamma(\tau) I_{i,t-1} + x'_{i,t}\beta(\tau). \tag{3.2}$$

By definition, the τ-th quantile of the distribution of democracy is the value $Q_D(\tau)$ such that $Pr(D \leq Q_D(\tau)) = \tau$. The quantile τ corresponds to the area under the unconditional distribution of democracy, bounded by zero on the left and the value Q_D on the right. Thus, a model estimated at $\tau = 0.5$ will produce evidence on the effect of income at the median of the democracy distribution. This model, more commonly known as a Laplace median regression, is often contrasted with the mean regression discussed above. However, our interest lies in estimating the conditional quantile functions at all quantiles of the distribution of democracy, paying particular attention to the estimation of the relationship in the tails of the distribution, that is where τ is close to either 0 or 1.

We will restrict our attention to a linear specification of the conditional quantile functions. This model can be estimated by solving,

$$\arg\min \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(D_{i,t} - \gamma(\tau) I_{i,t-1} - x'_{i,t}\beta(\tau)), \tag{3.3}$$

using interior point methods. The piecewise linear quantile loss function $\rho_\tau(u)$ is defined as $\rho_\tau(u) = (\tau \mathbf{1}\{u > 0\} + (1-\tau)\mathbf{1}\{u < 0\})|u|$ following Koenker and Bassett (1978). It can be shown that $(\gamma(\tau), \beta'(\tau))'$ minimize $E[\rho_\tau(D_{i,t} - \gamma(\tau) I_{i,t-1} - x'_{i,t}\beta(\tau))]$. The above model will be referred to as the pooled quantile regression

model, for although it is applied to panel data it does not estimate individual specific effects and further extensions which will be introduced below.

We focus on quantile regression as an attempt to capture the underlying observed heterogeneity in the relationship between democracy and its determinants such as income or education. Hence we estimate the relationship at different quantiles of the conditional distribution of democracy for $\tau \in \{0.1,\, 0,25,\, 0.5,\, 0.75,\, 0.9\}$. This design measures the effect at each of the three quartiles and also at the first and last decile. This allows us to gain a comprehensive view on how the relationship changes with the distribution of democracy.

It is important to note that quantile regression is not the same as mean regression applied to different subsets of the data ordered by the distribution of the dependent variable. Thus, while we estimate the quantile regression function at the low quantile $\tau = 0.1$ in order to ascertain the extent to which income and education condition democracy in the lower tail of the distribution of democracy, this is very different from estimating a mean regression where we condition on data in the low tail of the distribution. Thus, $Q_{D_{i,t}}(0.1|I_{i,t-1}, x_{i,t})$ is not the same as $E(D_{i,t}|D_{i,t} < c, I_{i,t-1}, x_{i,t})$, for some appropriately chosen c meant to capture the lower tail of the distribution. In particular it may be the case that the above moment does not exist. Furthermore, there is no theory which tells us how to choose or interpret the parameter c while τ has a natural interpretation. Quantile regression captures the effect of the covariates at a particular quantile of the distribution of the dependent variable, whereas the above suggested truncated mean regression estimates the conditional mean in a subsample of the data, ignoring the rest of the distribution of the dependent variable.

In order to facilitate comparisons across model specifications and econometric procedures we present results side-by-side in Tables 1, 2 and 3. Since for each specification quantile regression delivers not one set of estimates but five for each of the τ quantiles used, the tables can be interpreted in the following way. Consider the models presented in Table 1. The first model consists of the baseline pooled quantile regression described above. The model regresses the Polity Measure of democracy on the one period lagged log GDP per capita and the log population in the current period. The estimated coefficients at each of the quantiles are given in the first five columns labeled by the corresponding quantiles τ. The last column labeled "Mean" presents the estimated coefficients of a standard mean regression most closely associated with the quantile regression procedure employed in the corresponding quantile model. Thus, for the pooled quantile regression setup this is just OLS on the entire sample. As we shall see below however, once we take the panel data structure into account we shall employ other procedures such as fixed effects or instrumental variables for both the quantile regression and the mean regression. Given the fact that quantile regression produces five sets of coefficient estimates for every model it is unavoidable to have several related tables which the careful reader will need to navigate. In addition to containing information on different model specifications each table duplicates the

results for each of our two variables for the democracy, the Polity Measure and the Freedom House Measure.

As is customary we report standard errors in parenthesis. These were obtained using the bootstrap.[5] The general procedure involves sampling pairs formed of the dependent variable and the set of independent variables with replacement to accommodate different forms of heteroscedasticity. In our panel setting, a similar strategy is to first sample pairs of observations with replacement, and then randomly sample within groups either with or without replacement. The first strategy provides a reliable approximation for the precision of the estimates when T is relatively large (see, e.g., Davison and Hinkley 1997). The empirical covariance matrix can be computed given B bootstrap estimates of the coefficients. Due to the already dense tables required we shall not report significance for each quantile coefficient, although the reader can easily compute the t-statistic in each cell.

3.1.1. *An Inverted U-shaped Relationship between Income and Democracy?*

We shall now turn our attention to the evidence derived from the use of quantile regression as introduced above in uncovering the relationship between income and democracy. The pooled quantile regression model for the Polity measure estimates an inverted U-shaped relationship between income and democracy over the quantiles of the democracy distribution. Thus in the left tail of the democracy distribution we estimate a coefficient of 0.006 at the $\tau = 0.1$ quantile. The effect increases to 0.262 at the median of the democracy distribution and declines to 0.071 in the right tail of the distribution for $\tau = 0.9$. By contrast the coefficient on lagged GDP per capita is 0.170 if the model is estimated by OLS.

Similarly, if we estimate the same pooled quantile regression model for the Freedom House measure we estimate a coefficient of 0.073 at the $\tau = 0.1$ quantile, 0.252 at the median quantile and 0.145 at the $\tau = 0.9$ quantile. The mean effect of log GDP on democracy is 0.183.[6]

These results suggest that a 10% increase in log GDP per capita increases the democracy score of a country in the lower tail of the democracy distribution by less than 0.01. Recall that the difference between the mean democracy scores of Western countries to that of Sub-saharan countries is over 0.6. Thus, income appears to have a negligible effect on improving democracy in countries with low democracy scores. On the other hand the impact of income on democracy is three to four times larger in countries situated at the median

[5]The bootstrap and alternative resampling methods for cross sectional quantile regression has been investigated, among others, by Buchinsky (1995), Hahn (1995), Horowitz (1998).

[6]Note that our estimates of the mean effect are higher than the corresponding values reported in Acemoglu et. al. (2008). This is due to the different model specification. We do not include a lagged dependent variable in our specifications for reasons that will be discussed in Section 5.4.

of the democracy distribution. In practical terms this means that a 10% rise in income is likely to have a much larger impact in Latin America than in Sub-saharan Africa. The results lean towards providing support for the Epstein et. al. (2006) view that income is a much more potent engine of change in "hybrid democracies". In order to verify that this is indeed a quantile effect and not driven simply by the presence of countries with more unusual characteristics at specific locations within the distribution of democracy we also ran a number of specifications where we excluded certain countries which may be driving this effect. Thus, we excluded countries in Eastern Europe and the former Soviet Union, certain African countries or Muslim countries. In every case however we obtained a similar inverse U-shape pattern of the estimated coefficients even though the composition of the distribution changed. This indicates that the results do indeed vary over the distribution of democracy and are not driven by the inclusion of certain groups of countries.

In Table 2 we extend the model by adding a schooling variable corresponding to the percentage of the population in a given year with a secondary education. This reduces the effect of income on democracy at all quantiles as one would expect given the correlation between income and schooling. In the regressions using the Polity measure we also estimate an inverse U-shaped relationship between schooling and democracy over the quantiles. The estimated effect increases from 0.015 at the $\tau = 0.1$ quantile to 0.049 at the median and decreases to 0.001 at the $\tau = 0.9$ quantile. Similar results are obtained for the Freedom House measure.

The inverse U-shaped relationship was obtained from a pooled quantile regression model. We ought to be concerned, however, that the estimated relationship is not econometrically robust due to the lack of exogenous variation in income and schooling or due to ignored country specific heterogeneity. These issues will be explored in much greater depth in the next sections where we extend the quantile regression model to take into account endogeneity and different specifications of country effects. Without going into extensive details, Table 1 and Table 2 also include quantile regression results using instrumental variables and country-specific effects.

In the next section we discuss different choices of instruments and their relative merits. In Tables 1 and 2 we use the IV Set 1 which consists of geographic variables traditionally associated with economic development: mountainous terrain, geographic latitude and the distance to the nearest port. Using these instruments to account for the potential endogeneity of GDP in the model we estimate the effect of income on democracy at the previous quantiles. For both measures of democracy, we find that qualitatively the results are very similar. The estimated inverse U shape remains and is in fact more pronounced. Thus, the effect of income on democracy falls in the low and high quantiles and is higher at the median. For the Polity, measure the estimated coefficient on the $\tau = 0.1$ quantile changes from 0.06 to 0.02, while for the median it increases from 0.262 to 0.369. The estimated coefficient on the $\tau = 0.9$ quantile increases slightly from 0.071 to 0.088.

A similar pattern is observed for the Freedom House measure and for the model that includes schooling.

In order to account for unobserved heterogeneity at the country level we can augment the model by including country specific effects. As we will discuss in Section 5, in a panel data quantile regression model country effects may imply two different model specifications depending on whether they act as a location shift or distributional shift. The results in Table 1 correspond to a location shift model. We notice immediately that the inverse U-shape configuration of the estimated effects disappears altogether. In fact, for the Polity measure the estimated coefficients are 0.001 at the $\tau = 0.1$ quantile, -0.003 at the median and -0.073 at the $\tau = 0.9$ quantile. Once we add country effects the estimated relationship between income and democracy disappears at all quantiles.

Although our discussion so far has been focused on point estimates it is important to remark that these results cannot be explained away by statistical significance. For practical considerations we do not discuss the significance of the estimates of the coefficients at every quantile. However, it is easy to see that the measured inverse U-shape relationship is statistically significant for both the pooled quantile regression model and the IV model as the estimated standard errors are very small at every quantile. The standard errors increase substantially in the fixed effects model implying that the estimated coefficients on income are either zero or statistically insignificant. Similar results hold for the estimated coefficients on the schooling variable.

4. **Instrumental Variables**

In the previous section, we have briefly discussed the notion that conditional quantile function of Equation 3.2 may not capture the desired structural relationship if the sampling of income across the sample is independent of the error term. Thus, both income and democracy may be manifestations of some other latent variable not expressed in the conditional quantile function. This induces non-random sampling according to the relationship,

$$I_{i,t} = \delta(x_{i,t}, z_{i,t}, v_{i,t}), \tag{4.1}$$

where $z_{i,t}$ is a vector of instrumental variables independent of the structural disturbance but related to income and $v_{i,t}$ is an additional error term correlated with the disturbance of the democracy equation. Thus, we relax the assumption that income and the individual disturbance are uncorrelated. The idea is familiar from classical linear econometric techniques where it is known that it leads to biased estimates. Similarly, in the quantile regression setting the biased sampling induces an endogeneity bias in the estimated coefficients. This problem can be solved with the aid of additional exogenous variables, known as instruments, which are independent of the unoberved disturbances but structurally related to income by the above reduced form

relationship. We follow the method proposed in Chernozhukov and Hansen (2008).

As before we also assume that the conditional quantile relationship is monotonic in τ. The objective function for the conditional instrumental quantile relationship is given by:

$$R(\tau,\gamma,\beta,\mu) = E\left[\rho_\tau\left(D_{i,t} - \gamma(\tau)I_{i,t-1} - x'_{i,t}\beta(\tau) - \mu(\tau)\hat{I}_{i,t-1}\right)\right], \tag{4.2}$$

where $\hat{I}$ is obtained as the linear projection of the exogenous variables x and z on the endogenous variable. The estimation procedure of Chernozhukov and Hansen (2008) proceeds in two steps. First we minimize the objective function above for β and μ as functions of τ and γ,

$$\left(\hat{\beta}(\gamma,\tau),\hat{\mu}(\gamma,\tau)\right) = \text{argmin}_{\beta,\mu}\ R(\tau,\gamma,\beta,\mu). \tag{4.3}$$

Then we estimate the coefficient on the endogenous income variable by finding the value of γ, which minimizes a weighted distance function defined on μ:

$$\hat{\gamma}(\tau) = \text{argmin}_{\gamma}\ \hat{\mu}(\gamma,\tau)'\hat{\Omega}(\gamma)\hat{\mu}(\gamma,\tau), \tag{4.4}$$

where $\hat{\Omega}(\gamma)$ is the inverse covariance matrix of $\sqrt{NT}(\hat{\mu}(\gamma,\tau) - \mu(\gamma,\tau))$. Notice that the expression above implies that the procedure effectively minimizes the Wald statistic of the test $\mu(\gamma,\tau) = 0$. So far we have only discussed the procedure in terms of income as the endogenous variable, but it also applies to specifications which include education as an endogenous variable.

As Robinson (2006) remarks, Acemoglu et. al. (2008) are the first ones to propose an instrumental variable approach to the identification of the causal effect of income on democracy. Since their proposed instruments are not uncontroversial we shall explore some alternatives below.

4.1. **Geographic Instruments**

While the political science literature on democratization seems to have ignored the potential endogeneity of income in this specification, economics has traditionally stressed differences in geography as a potential determinant of economic development (Acemoglu, Johnson and Robinson, 2005). Geography is thought to structure the opportunities for material welfare experienced by economic agents. This is particularly salient in agrarian societies which are heavily dependent upon climate and geography as technological constraints. Acemoglu, Johnson and Robinson (2001) also suggest health as a channel through which geography influences economic development. Many diseases like malaria are only found in certain areas of the world with devastating effects on economic development.

We use the log of mountainous terrain, geographic latitude and the log of air distance to the nearest port as our instrumental variables set I. This set of instruments was used to derive the results described above. It

is designed to be a first approach to dealing with endogeneity within the context of this model.

While these instruments are arguably correlated with GDP per capita, the correlation might be weak. Furthermore, if we are prepared to assume that geography determines the economic choice set available to an individual it is easy to extrapolate that it also shapes the political landscape by influencing political preferences. Thus, geography might not provide us with adequate instruments after all.

4.2. **Trade-Weighted World Income Instrument**

Acemoglu et. al. (2008) use two different instruments for income in their linear specification. The first instrument corresponds to the savings rate in the previous five-year period. While we agree that it is difficult to find a priori reasons why the savings rate would affect democracy, it is nevertheless conceivable that the two variables are correlated, especially in developing countries. Hence we only use the second instrument, the trade-weighted world income. Let $\omega_{i,j}$ denote the trade share between country i and j in the GDP of country i between 1980-1989. Then we can write the income of country i at time $t-1$ as,

$$I_{i,t-1} = \zeta \sum_{j=1, j\neq i}^{N} \omega_{i,j} I_{j,t-1} + \epsilon_{i,t-1} \tag{4.5}$$

where income is measured as log of total income. Acemoglu et. al. (2008) suggest the use of the weighted sum of world income for each country, $\hat{I}_{i,t-1}$ as an instrument. Notice that the weights are not estimated but correspond to the actual trade weights. This instrument may be problematic if income in country j, $I_{j,t-1}$, is correlated with democracy in country j which itself is then correlated to democracy in country i. Furthermore, the trade weights, $\omega_{i,j}$ may be correlated with the relative democracy scores of countries i and j. We know that more democratic countries tend to be more open to trade. Furthermore, there may be political economy reasons why countries prefer to trade with countries which have similar political regimes. In this sense it might not be very surprising that Iran and Cuba have preferential trade agreements in place.

4.3. **Global Economic Factors**

The world income instrument described above has an appealing interpretation since it is designed to capture the intuition that business cycles are to some extent correlated with events in world markets. Trade reflects only one aspect of this international dimension. Recently, Harding (2008) argued that the stochastic dimension of an economy is potentially rather large and thus numerous global factors that have a substantial impact on a given economy could be identified. A statistical factor model can be employed to recover a set of orthogonal factors that can act as international sources of domestic economic fluctuations (Kose et. al., 2003). These global factors drive, to some extent, the domestic business cycle independently of the political regime of a country.

Econometrically, we can write the following linear factor model for income:

$$I_{t-1} = \Lambda F_{t-1} + U_{t-1}, \tag{4.6}$$

where I_{t-1} is the observed N dimensional vector of log GDP, F_{t-1} is a p-dimensional vector of global factors and U_{t-1} is a vector of idiosyncratic errors. The coefficient matrix Λ is a matrix of individual specific weights (factor loadings). Since only log GDP is observed we need to use a statistical procedure such as Principal Components Analysis (PCA) applied to the covariance matrix $\Sigma = (1/T)I_{t-1}I'_{t-1}$ to recover the latent factors $\hat{F}_{t-1}$.

By construction, this method separates the commonalities F_{t-1} from the idiosyncratic shocks U_{t-1}. In order to further exclude the possibility that the political regime affects the global factors through its effect on income we construct different values of the instruments for each country by excluding the country from the analysis. Thus, the instruments for country j correspond to the global factors estimated from the matrix of GDP measures for all the countries in the world except country j.[7]

In many financial applications it is of interest to know the exact number of factors (Bai and Ng, 2002; Harding, 2008; Stock and Watson, 2002). This issue is substantially complicated by the fact that for persistent time series such as GDP, the latent factors often have a dynamic structure. If this fact is not accounted for a plain application of PCA will recover both the factors and their lags under certain identification criteria on their strength relative to the variance of the noise (Harding, 2007). In the present case, we can ignore this debate since we only employ the estimated factors as instruments. The estimated factors are not directly interpretable since they correspond to combinations of many different economic fundamentals operating at the global level. Nevertheless, they are valid instruments and can be employed in the instrumental variable specification of our model.

Basic statistics of the instruments are presented in Table 3, in the online appendix. While it is not necessary nor feasible to interpret the proposed instruments in a concrete economic setting, it is interesting to note that the chosen principal components have similar statistical properties across world regions. The first factor appears to be more important for Western democracies, while the second factor for Sub-saharan African countries. Similarly the fourth factor appears less important for Eastern Europe and the Soviet Union. This is consistent with the notion that we are capturing global economic factors acting as sources for the transmission of international business cycles while still expecting regional variation in their impact.

4.4. **Robustness to Instrument Choice**

In Table 3 and Figure 3 we present the estimation results for the baseline pooled model after we instrument for income with one of our three sets of instruments. The left panels in Figure 3 compare the estimated

[7]We have also experimented with excluding all countries in the geographic region to which j belongs in the construction of the instrument for j but have obtained very similar results.

quantile effects of income on democracy between the quantile regression (QR) estimator corresponding to the pooled model and the instrumental variable (IV S1-3) estimators corresponding to the instrumental variable quantile regression estimators using the three sets of instruments. In Table 3 we report results for $\tau = \{0.1, 0.25, 0.5, 0.75, 0.9\}$. Several of the estimated coefficients at the 0.9 quantile, however, are zero, which makes us doubt that the instrumental variable estimator worked for this quantile. For computational robustness, we restrict our attention in Figure 3 to twelve quantiles between 0.2 and 0.8. We believe that some of the other estimated nil effects can also be explained as computational limitations. However, results appear fairly robust, at least from a computational point of view between the 0.2 and 0.8 quantiles.

For both measures of democracy, estimation using the first set of instruments, corresponding to the geographic variables, only amplifies the inverted U-shaped relationship as described in the previous section. The second set, corresponding to the trade weighted world income instrument, only replicates these results for the lower half of the distribution. The estimated coefficients for the high quartiles are higher than at the median. For the Polity measure the estimated coefficient increases from 0.261 to 0.302 while for the Freedom House measure it increases from 0.136 to 0.170 at the $\tau = 0.75$ quantile.

If we now employ our last set of proposed instruments, we find that the inverted U-shape relationship is preserved for the Polity measure but not for the Freedom House measure. The estimates for the Polity measure are slightly lower than those for the pooled quantile model and almost 50% smaller than those obtained using the first set of geographic instruments. The corresponding relationship for the Freedom House measure, however, estimated a very high impact of income at high quantiles of democracy. Overall, it predicts an increasing relationship between income and democracy over the quantiles.

While we have obtained mixed results under the different instrumentation strategies, it appears that most specifications preserve the inverse U-shaped relationship between income and democracy at different quantiles of democracy. The different specifications appear to disagree mostly on the effect of income in the high quantiles. However all specifications, appear to confirm a positive effect of income on democracy in the middle range of the democracy distribution. All but one specification estimate a lower effect of income on democracy in the low quantiles compared to the median. Thus, after implementing three different strategies addressing the potential endogeneity of income, we still find higher income to be a more important force for democratization in "hybrid regimes" close to the median of the democracy distribution and to have almost no effect in non-democratic regimes.

5. **Panel Data**

We have attempted to estimate the structural effect of income on democracy using a number of instrumental variables strategies. Acemoglu et. al. (2008) show, in the context of linear regression, that this might be insufficient in the presence of long run unobserved institutional factors which condition the equilibrium paths

of both economic and political development. This challenges the notion that a small set of contemporary explanatory variables such as income or population can do a good job at explaining the distribution of democracy around the world. As we anticipated in Section 3.1.1, if we augment our econometric specification with country specific effects, we are going to find that the inverted U-shaped relationship breaks down. In this section we will explore this effect in greater detail.

First however, let us provide some preliminary evidence on the persistence of the bimodal distribution of income even after conditioning on income and schooling. To this effect, we construct the (kernel smoothed) conditional density of democracy for each of the two measures of democracy. The two conditioning variables are income and schooling. In Figure 4 we compare the unconditional and conditional distributions for the two measures of democracy and for three partitions of the conditioning variables, "low," "median" and "high." These partitions are chosen based on the quantiles of the respective variables. Thus, when we compute $f(D|Low\ I)$, we compute the distribution of democracy conditional on income being below the 25th percentile. Similarly, "high" corresponds to the conditioning variable being above the 75th percentile with the remaining mass allocated to the "median."

As can be seen from Figure 4, in almost all cases, the conditional distribution of democracy retains its bimodal shape after conditioning on income and schooling. Income and schooling are only partially able to explain the observed shape. Consider for example the distribution of the Polity measure in the first row of Figure 4. Notice the change in the distribution of democracy as we first condition on low income. The unconditional distribution has the bimodal shape discussed above. After conditioning we find the frequency of country-years with low democracy scores increasing while the frequency of country-years with high democracy scores decreasing. The bimodal shape remains, illustrating the fact that even after conditioning on low income we still find countries with both high and low scores of democracy. Similarly, when we condition on high income we find a higher frequency of country-years with a high democracy score but also a substantial number with low democracy scores. This pattern is repeated throughout the panels in Figure 4.

This points to the possible existence of more long run influences that shape the political development of a country and which is not susceptible to the contemporaneous effects of income and schooling but itself may determine these. As previously hinted, we will attempt to use panel data methods to control for these forces.

5.1. **Location Shifts**

The simplest possible quantile model which allows for fixed effects is given by the following equation for the τ-th conditional quantile functions of democracy for country i at time t,

$$Q_{D_{i,t}}(\tau|I_{i,t-1}, x_{i,t}, \mu_i) = \gamma(\tau)I_{i,t-1} + x'_{i,t}\beta(\tau) + \mu_i. \tag{5.1}$$

This extends the standard pooled model discussed above by allowing for an individual specific effect μ_i, which does not vary across quantiles. These effects should be simply interpreted as country-specific intercepts that shift the conditional quantiles functions by μ at each quantile. We call this model the location-shift quantile regression model (labeled FEQR (LS) in the tables and graphs). This model can be estimated for J quantiles simultaneously by solving,

$$\arg\min \sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{t=1}^{T} \omega_j \rho_{\tau_j}(D_{i,t} - \gamma(\tau_j) I_{i,t-1} - x'_{i,t}\beta(\tau_j) - \mu_i), \tag{5.2}$$

using the method described in Koenker (2004). The weights ω_j control the influence of the j-th quantile on the estimation of the quantile effects. In the present study, we will employ equal weights $1/J$.

If we now turn our attention to Figure 3, we can evaluate the plots of the quantile effect of income on democracy at different quantiles after we estimate individual specific effects. (The interested reader can see the online appendix for a similar analysis based on schooling). The figure reveals that once we add country effects the inverted U-shaped relationship disappears. This result is consistent with the mean regression results of Acemoglu et. al. (2008), who also find that once country effects are added to the specification the relationship between income and democracy disappears. Within the context of quantile regression however, we can take the analysis a step further and investigate the nature of the country effects. If we associate different political outcomes with different sets of long run institutions, we can ask the question whether these institutions have the same impact at different quantiles of the democracy distributions.

5.2. **Distributional Shifts**

In large T panel data it is possible to estimate a different value of the individual effect for each quantile of the conditional distribution of the response. This is a novel opportunity which allows us to evaluate the role of the country effect, and by implication the long run drivers of democracy which it proxies for, as they vary over the distribution of democracy. To our knowledge this is the first time that distributional country effects have been employed in empirical work. To introduce the concept, let us first estimate a quantile group effect before proceeding with the estimation of country specific effects.

The model of interest for the τ-th conditional quantile function of democracy for country i at time t in the presence of group effects is,

$$Q_{D_{i,g,t}}(\tau | I_{i,g,t-1}, x_{i,g,t}, \mu_{i,g}) = \gamma(\tau) I_{i,g,t-1} + x'_{i,g,t}\beta(\tau) + \mu_{i,g}(\tau). \tag{5.3}$$

We can allocate each country to a group g based on the quartiles of the distribution of the mean value of a country's variable of interest. Thus, for example we can divide the countries into four groups based on average schooling, "low," "middle-low," "middle-high," and "high." A similar grouping can be done for income. We also consider two variables meant to capture the presence of long run institutions based on the

work of Acemoglu et. al. (2008). These correspond to estimated settler mortality and population density in 1500. In Tables 5 and 6, in the online appendix, we present the composition of each of the four groups for the four variables of interest for all countries for which data was available.

In Figure 5, we present the plots of the estimated quantile group effects for the first three groups using the "high" group as reference point. In each graph, the continuous dotted line shows the point estimates, and the shaded region represents a 95% confidence interval for the point estimates. The estimate group effect for the "low" schooling group is negative but has a positive slope over the quantiles of the democracy distribution. The group effect for the "middle-low" group is also negative but closer to zero than the effect for the "low" group. Both group effects are statistically significant at all quantiles. The estimated group effects for the "middle-high" schooling group are very close to zero and statistically insignificant. This shows that the presence of individual effects is much more pronounced for low schooling countries and that the effect of these long run factors diminishes for more democratic countries. This shows that there is substantial heterogeneity impacting country specific effects on democracy.

If we now group the countries by income we find a similar pattern to the one where we group the countries by schooling. The effect is negative and significant only for low income countries in the lower quantiles of the democracy distribution. Additionally we find positive significant quantile group effects for countries with "low" and "middle-low" settler mortality but no significant group effects for countries with different population densities in 1500.

This analysis strongly suggests that country specific effects are not likely to be uniformly important for all countries in the sample. Moreover their significance diminishes as countries become more democratic. The impact of country heterogeneity is likely to be very mixed. The impact of long run factors is likely to be stronger in countries with both a challenging past and unfortunate present economic and historical circumstances, but that once a country overcomes the effect of bad initial conditions these long run factors will diminish in importance.

In order to explore this point further, we can now proceed to estimate individual quantile effects for all countries. The τ-th conditional quantile function of democracy for country i at time t in the presence of individual effects is,

$$Q_{D_{i,t}}(\tau|I_{i,t-1}, x_{i,t}, \mu_i) = \gamma(\tau)I_{i,t-1} + x'_{i,t}\beta(\tau) + \mu_i(\tau). \tag{5.4}$$

We call this model the distributional-shift country effects quantile regression model (labeled FEQR (DS) in the figures). Notice that we now estimate a series of individual effects, one for each quantile, for every country. The procedure requires the estimation of N country specific effects and thus is only feasible for large T. For identification purposes, we estimate a model without an intercept term. This model can also easily be extended to the instrumental variable case for each of the three sets of instruments employed in this paper (labelled IVFE-S1 through IVFE-S3).

In Figure 3, we compare the estimated coefficients for income across quantiles with the results we obtained from our previous models. In most specifications, the estimated coefficients on income and schooling are close or identical to zero. The lack of a statistical relationship is particularly emphasized at low quantiles for $\tau < 0.5$. Depending on the chosen specification and the instruments used, we find some variation at the high quantiles. However, this is not very robust and the estimated coefficients are close to zero even though they can take both negative and positive values at the high quantiles. Results appear consistent across the different sets of instruments. Results based on the first set of geographic instruments are rather different from the results employing the other sets of instruments. It appears to find a downward sloping relationship between income and democracy over the different quantiles of the distribution. The other two sets of instruments appear to predict almost no effect of income on democracy over the quantiles.

Allowing the individual effects μ_i's to have a distributional shift does not appear to change the estimated coefficients on the right hand side variable but does potentially provide us with more information on the heterogeneous impact of the long run country specific factors they are designed to capture.

In Figures 6, we plot the country effects estimates for the models of Polity measure of democracy. The results for the Freedom House democracy score were similar and therefore can be found in the online appendix. We only plot the results for countries with at least one individual effect significant the at 1% level. All effects are however estimated and point-wise confidence intervals are computed at each quantile

To our knowledge this is the first time such plots have been employed. Thus, there is no established tradition in interpreting them. Quantile country effects, which are allowed to vary over the distribution of the dependent variable, are estimated at all quantiles. This is the case even when a particular country does not vary in its position in the distribution of democracy over time. For example we estimate a quantile country effect which varies with τ for the US even though the US scored at the very top of both democracy measures throughout the sample. This feature is rather counterintuitive even though it appears to be mathematically correct. It is challenging to interpret what exactly is identified in a situation where the dependent variable does not change over time for one country. Mathematically, the quantity is identified as long as income and population change over time. Notice that the objective function for quantile regression involves minimization over terms in $\rho_{\tau_j}(D_{i,t} - \gamma(\tau)I_{i,t-1} - x'_{i,t}\beta(\tau) - \mu_i(\tau))$. Thus, even though $D_{i,t}$ is constant over time, the term $D_{i,t} - \gamma(\tau)I_{i,t-1} - x'_{i,t}\beta(\tau)$ is not. Moreover, notice that $\gamma(\tau)$ and $\beta(\tau)$ are constrained to be the same across countries i while $\mu_i(\tau)$ alone varies across countries. For low quantiles, we have seen before that the effect of income (and other regressors such as population) is likely to be close to zero. Thus, the estimation algorithm will try and "match" $\mu_i(\tau)$ with $D_{i,t}$. Thus, that τ-th quantile of democracy should be similar to the country effect μ_i. Because we estimate a model without an intercept for identification, we thus expect the estimated quantile country effect to be close to 1 for countries such as the US which don't vary over time and for which we estimate the quantile effect at low quantiles. This is indeed the pattern that we observe

in Figure 6. The evidence in Figure 6 on individual effects is thus consistent with the claim that income is particularly ineffective at explaining political outcomes in the lower quantiles of the democracy distribution.

In all cases however, the effect of country specific effects diminishes and is estimated with wider error margins at high quantiles. While we were not able to obtain robust estimates across different econometric specifications of the effects of income and schooling at high quantiles, our earlier results point to the presence of some effect. We believe that this is due to additional heterogeneity in the way income and schooling affect countries at the same quantile. Intuitively, this suggests that income and education play a very different role in countries with similar levels of political development which are not adequately captured by fitting a common structural equation for the quantile. Econometrically, it is not possible within the confines of our data to estimate a random coefficients model that would allow for additional heterogeneity though country specific slope parameters at each quantile. While additional elaboration of this point is beyond the scope of our paper, we should remark that a vast literature exists in political science on the institutional differences between democracies with similar ranks on the democracy scales employed in the analysis. Lijphart (1999) documents the institutional heterogeneity in 36 democracies by means of ten variables arguing for a split between "majoritarian" and "consensus" democracies based on the extent to which political power is concentrated or shared. These democracies tend to be rather different in their approach to the management of the economy or the welfare state. If Lijphart (1999) is correct, then income and schooling have very different effects in democracies even though superficially these democracies appear rather similar in our rankings. Similarly, the literature on "varieties of capitalism" (Hall and Soskice, 2001) documents a wide range of differences in the political economies of developed nations. To conclude, we believe that the evidence in this paper points to a diminished impact of long run country specific effects for observations in the upper tail of the democracy distribution and to additional heterogeneity in the effect of income and schooling for these observations. And hence, while our data does not allow us to study these additional differences, we do not want to diminish its importance in future research.

We can compare the estimated individual effects across countries at a given quantile. This provides us with additional information on the heterogeneous impact of long run factors at different quantiles. In Figures 7 and 8, we plot the estimated quantile country effects for the 0.3 and 0.7 quantiles, $\hat{\mu}_i(0.3)$ and $\hat{\mu}_i(0.7)$, against the corresponding mean values for democracy for each country. Plots against schooling and income can be found in the online appendix. Furthermore, in each figure we report six scatter plots, where for ease of interpretation, we have grouped the countries into six regions.

First, let us notice that the figures in the online appendix reveal the estimated quantile country effects to be uncorrelated with income or schooling in the sense that within each region countries tend to have comparable levels of income and schooling but very different estimated country specific effects. Figures 7 and 8 show a high degree of correlation between the country-specific mean democracy score as measured by the Polity database and the estimated quantile effects. This is to be expected since, as we have previously

seen, in many cases the country specific effects capture most of the variation. In particular, notice how this correlation is very high at the 0.3 quantile, since in the left tail of the democracy distribution the quantile effect of income and schooling is close to zero. For the 0.7 quantile, however, the correlation is less perfect due to the effect of the structural component of the quantile function. While the results do not appear to be very robust, income, schooling and population play a role here, making the country specific effects less important. Furthermore, note the different scales for the estimated country effects $\hat{\mu}_i(0.3)$ and $\hat{\mu}_i(0.7)$. The estimate $\hat{\mu}_i(0.3)$ ranges between 0 and 1, which is consistent with the observation that when the explanatory variables play almost no role in the quantile function, the country specific quantile effects will explain all the variation in the dependent variable. If we now look at the range of the quantile effect $\hat{\mu}_i(0.7)$, we notice that they range between -0.6 and 0.6. Since our democracy scores range between 0 and 1 this appears to suggest that at most 40% of the democracy score may be captured by the included variables while the rest is due to the country specific quantile effects.

5.3. **Further Robustness Checks**

In this paper, we have employed a wide array of quantile estimation techniques which provide a thorough exploration of the degree to which our results are robust to different econometric specifications. The results presented in this paper focus on three main explanatory variables: income, schooling and population. An earlier version of this paper also explores additional variables such as economic growth, oil production, ethnicity or religion. We have not found these to affect the results substantively and for reasons of conciseness we are not reporting them here.

There are a couple of important differences in the way the data was employed in our paper compared to Acemoglu et. al. (2008). First, we use annual data rather than 5-year panels. We replicate some of the basic regression estimates for the pooled, instrumental variables and country effects models using the 5-year panel data of Acemoglu et. al. (2008). We obtain very similar coefficient estimates across quantiles. In particular we also estimate an inverted U-shaped relationship between income and democracy for the pooled and instrumental variables models, which disappears as soon as individual effects are added to the specification. We chose to estimate the model on annual data even though democracy is a slowly moving process in order to be able to estimate quantile country effects more precisely. The resulting point estimates however are very similar. These results are not reported in this version of the paper, but are available for the interested reader in the online appendix.

However, we do depart from the approach of Acemoglu et. al. (2008) in a crucial way. We do not estimate a dynamic model, i.e., we do not include a lagged measure of democracy on the right hand side of our specification. The inclusion of a lagged dependent variable in quantile regression models has not been explored so far in the econometric literature. Lacking a theoretical econometric foundation to deal with dynamics in this context, we decided to avoid this issue in the present paper. There is no immediate way

of implementing a dynamic setting as we cannot separately identify the country idiosyncratic effect. We do conduct one additional robustness check, however. In the online appendix, we provide additional evidence on the distributional effect of income on democracy by limiting the sample to countries with positive standard deviation of democracy. If an estimator for dynamic panel data would be available for quantile regression then it could at most be applied to this subset. We find very similar results for this subset of the data. In particular notice the inverted U-shaped dependence for the pooled and instrumental variable estimators which disappears once we include country effects. In any case, the inclusion of a lagged dependent variable is not likely to increase the chances of finding income or schooling as significant driving forces for democracy in our model. By analogy, in mean regression settings, lagged dependent variables absorb much of the time variation especially for slow moving processes such as democracy.

6. **Patterns of Democracy**

By focusing on the scatter plots for the quantile country effects in Figures 7 and 8, as well as Figures 6 to 9 in the online appendix, we can learn more about the patterns of political development in the world. While some variation is explained by the included variables such as income, schooling and population, their effect appears not to be robust or consistent across most econometric specifications. The scatter plots for the country effects may provide some guidance to the unexplained variation in democratic and authoritarian rule across the world.

Western Democracies and Japan corresponds to the region of consolidated modern democracies. However, even here we observe great diversity over the second half of the 20th century. While sharing similar levels of economic development, not all European countries have shared an equally democratic past over the last 50 years. Notice in particular how the estimated quantile country effects for Portugal and France are substantially lower at the 0.3 quantiles than at the 0.7 quantiles. This reflects the fact that at some point over the last 50 years these countries have experienced periods of authoritarian government. Portugal's first attempts at democracy was the establishment of a republic in 1910 that collapsed under economic pressures and led to the establishment of a right-wing military dictatorship by Antonio de Oliviera Salazar. Democracy was reintroduced in 1974 after another left-wing military coup. While France did not experience a dictatorship, its institutions changed drastically at the conclusion of the Algerian war when President Charles De Gaulle established the Fifth Republic in 1958. The constitutional reforms ended what was considered a Parliamentary system and established a semi-presidential system of government by strengthening the institution of the Presidency. Additionally, France experienced perhaps the most turbulent social protests of any European country in the late 1960s, as student strikes dominated Paris. The estimated quantile individual effects capture the idiosyncracies of these regimes in Europe after the end of World War II. By estimating different country effects for different quantiles we are implicitly allowing for time variation in the "fixed" effects. This challenges the association of these effects with long run development paths that jointly

determine both political and economic development. The evidence provided by the quantile country effects appears to underline the idiosyncracy of a country's development. The quantile effects are capturing this idiosyncracy as a mixture of long run factor and contemporary shocks not adequately controlled for by the included variables.

We can learn more about the patterns of political development in other countries in the world by focusing on the conditional quantiles of the model 5.4. To obtain the quantiles of democracy for each country, we evaluate the estimated model at the country-specific mean of log of GDP and log of population. Figure 9 presents the conditional and unconditional quantiles of democracy in several countries representing different regions of the world. In all the cases, we see that the vertical distance between the Polity measure of democracy and the prediction is relatively small, suggesting that the panel data model offers a satisfactory way of estimating the country-specific distribution of democracy. This figure also shows the relative limitation of the conditional mean approach, only providing acceptable predictions for the somewhat "degenerated" cases (e.g., US, and Botswana).

Greece and Spain reflect the fact that at some point these countries have experienced periods of non-democratic regimes. Their estimated democracy levels at the lowest quantiles are comparable to the less advanced democracies in the world, but the levels at the highest quantiles are comparable to the more established democracies. Greece emerged from the 19th century as an established monarchy which continued until 1967 when a military dictatorship arose. The junta faced popular opposition and international pressure and democratic elections were held in 1974. Spain, while now a constitutional monarchy with a responsive democratic government, was mired under the repressive regime of General Francisco Franco for a thirty-six year stretch from 1939 and 1975. Franco was a product of a nationalist movement which allied the fascist right with the anti-left middle class during the 1930's and then survived through support from Hitler and Mussolini even though officially neutral during World War II.

In sum, when considering the implication of our results for developed democracies, our model is able both to correctly predict the level of democracy and to explain a high level of diverse paths to consolidated democracy that Western European countries have taken. Most importantly, our estimated country-specific effects identify the set of South European countries, in addition to France, that has long occupied the attention of those studying democratization (Schmitter et al. 1986, Huntington 1993, Linz and Stepan 1996). However, at the same time we have also explained the political development in traditional liberal consolidated democracies, whose institutions changed very little over the past century. Our quantile country-specific effects approach is equivalent to counterfactual analysis, which allowed us to estimate the factors driving democracy even given the tight clustering of economically developed democracies in Western Europe.

Eastern Europe and the former USSR is the region of the world where present-day country-specific factors are most easily recognized as reflecting a legacy of communist institutions. This region has been especially important for students of democracy because of the stark divergence of institutional change following the

fall of communism. Contrary to original hopes of democratic leaders, the first decade of the post-communist experience in the region saw more countries slide back to authoritarianism than establish consolidated democracies. Our results explain this effect both through the nonlinear effects that income and education have, and through changing country-specific effects. In all our scatterplots of country-specific effects, the Czech Republic and the Baltic states cluster in the upper-right hand corners. Both according to the level of development and to the underlying cultural and historical legacies, these countries have been identified by scholars of the region as the best candidates for future democratic consolidation (McFaul and Stoner-Weiss 2004, Ekiert and Hanson, 2003). In contrast, former republics of the Soviet Union other than the Baltic states have had a remarkably difficult transitional period in the 1990s. Our analysis reveals this not only the effect of low income or educational attainment, both of which characterize states such as Kazakhstan and Uzbekistan, but also an effect of country-specific factors that in this case capture institutional legacies of the particularly oppressive form of communism that Stalin instituted following the occupation of these areas after the Second World War. Experts on the region have also identified the unique Soviet-style system of party elite rule to be an obstacle to building effective institutions of party competition (Jones-Luong 2002, 2003).

Our data also identifies Russia under Boris Yeltsin as falling on the opposite end of the spectrum from the developed democracies when it comes to country-specific factors. Because of the persistence of country-specific factors across different quantiles of democracy, Russia much like China represents a country where democratization is likely to encounter greatest obstacles. This finding is significant in that it corresponds to the turbulent institutional change in Russia not only in the 1990s but also in the period that followed, described by numerous studies of Yeltsin and Vladimir Putin's rule of the country (McFaul 2002a, McFaul 2002b, Fish 2005, Colton and Holmes 2007, Politkovskaya 2007, Shevtsova 2007a, 2007b, Colton 2008).Most of these authors ascribe the failure of democracy in Russia to a lack of political order and well-functioning institutions; but the problem that our analysis identifies is that even in the counterfactual scenario of having a developed democracy in Russia (better than 70 percent of the rest of the world's regimes), the country-specific factors that existed under the communists will persist. Another way of interpreting this finding is that changing country-specific factors such as historical legacies and the institutions of the state are unlikely to result in a consolidated democracy in Russia.

The Balkans illustrate perhaps the most interesting role of country-specific factors. When we allow country-specific factors to vary according to quantiles of income, we identify Albania and Romania forming a unique cluster, at the very opposite end from the Baltic countries discussed above. This result is consistent with the fact that the two countries experienced two of the bloodiest communist dictatorships in Europe. In Romania, despite its high growth rates in the period we studied, Nicolae Ceausescu's dictatorship destroyed entire segments of Romania's civil society, culture and state institutions. The brutality of Ceausescu's regime sparked the Europe's first bloody postcommunist revolution on December 22, 1998, when Ceausescu together

with his wife Elena was shot by a firing squad. The violent end of the communist dictatorship in Romania, however, did not lead to the establishment of democracy. As Grzymala-Busse's (2002) work in the region has uncovered, the systems of patronage and clientalism persisted throughout the 1990s, limiting the effectiveness of the party system in Romania and representing what our study has identified to be country-specific factors obstructing institutional change. Similarly, in Albania, until his death in 1985, Enver Hoxha ran a police state that not only suppressed dissent but also systematically dismantled the set of state and civil society institutions necessary for democracy. While Romania and Albania are extreme examples of the destructive communist legacies, weak civil society in particular has been an important obstacle to the quality of all new post-communist democracies (Howard 2003).

Even more so than in Albania and Romania, the most extreme case of destruction of democracy-supporting institutions and culture can be found in Yugoslavia. Dating back to its establishment in 1918, the country's changing authoritarian governments — starting with Serbian kings and ending with the Socialist Party — failed to resolve ongoing ethnic differences that culminated in bloody civil wars in Croatia (1991-1995), Bosnia (1992-1995) and Kosovo (1996-1999). What our analysis identifies as country-specific factors historians of the region such as Ivo Banac (1988) have explored in much greater depth. Banac (1998) explains why and how the question of nationalism emerged in Yugoslavia and how it was engineered and manipulated by ethnic leaders into a uniquely virulent strain of a destructive political ideology. The uniqueness of the Yugoslav case as a total dissolution of the state and a failure of institutional change is most starkly apparent in the volumes of court documents and eye-witness accounts that are now part of the historical record of the Hague Tribunal that indicted the country's ethic leaders for genocide and crimes against humanity (Cigar et al. 2002).

Sub-saharan Africa is a region of opposites and ranks as the world's poorest and least educated. Our results reveal that while the region is generally undemocratic, a few countries have established flourishing democracies. The quantile country effects scatter plots reveal a consistent clustering of countries at both the 0.3 and 0.7 quantiles. This is explained by the presence of long-run factors as opposed to idiosyncratic but transitory political shocks. Sierra Leone is one of the poorest countries in the entire world, ranking also among the worst countries in terms of life expectancy and education. Sierra Leone inherited democratic institutions after independence in the early 1960's, but political convolution and military coups have been affecting democracy in the country since the 1970's. Zimbabwe also inherited democratic institutions around the time of independence in 1980, but President Mugabe and his Zimbabwe African National Union - Pacific Front's (ZANU-PF) systematic violence to political opposition and human rights violations have been affecting the prospects of democracy. Sierra Leone's and Zimbabwe's patterns of democracy are consistent with the evidence presented in Figure 9.

On the other hand, countries such as Botswana and Gabon find themselves at opposite ends of the spectrum as far as their political development is concerned. Botswana took an unusual route to its current state a of

vibrant multi-party democracy that explains the almost invariant democracy levels across the quantiles in Figure 9. Freed from colonial rule in 1966, Botswana was one of the poorest countries in the world. But the discovery of diamonds in 1971 fueled the economy and generated extended network of government sponsored social programs which helped maintain the democratic process and encouraged peaceful transitions between its three presidents Seretse Khama, Ketumile Masire and Festus Mogae. By contrast, Gabon is also a very resource-rich nation with extensive mineral and oil deposits, yet its politics has been dominated for the past four decades by Omar Bongo, who runs a corrupt and repressive regime.

Our results in Africa therefore confirm the findings in the literature that emphasize the success story of democratic institutions in Botswana (Acemoglu, Johnson and Robinson 2003, Robinson and Parsons 2006), emphasizing that democracy is chiefly determined by country-specific effects and that political development in Africa cannot be described as emerging from development in the form of increasing income or education. Also in accordance with the literature that emphasizes the role of natural resources in the political economy of Africa (Collier and Hoeffler 2004, Fearon 2005). By taking into account the country-specific effects, our model was able to describe the contrasts that characterize the political development of Africa.

Latin America is a region associated with instability and diversity. Our results show that while some countries exhibit transitory shifts from democratic to non-democratic forms of government, others remain democratic most of the time. Chile has been one of the most stable democracies characterized by fair elections and respect for democratic institutions. However, one of the turning points in the political development of the region has been September 11, 1973, when Salvador Allende's socialist government was interrupted by a military intervention leaded by General Augusto Pinochet. This initiated a brutal regime that lasted more than 16 years. The repressive regime of General Pinochet is associated to a drop in the (normalized) Polity measure of democracy from 0.8 in 1973 to 0.15 in 1974. While having traditionally higher levels of economic development and education than its neighbor Chile, Argentina became an ungovernable, undemocratic society after 1930. President Juan Domingo Perón's administration that lasted 10 years (1945-1955) was characterized by centralized power and limited civil and political liberties. From 1955 to 1983, the government wavered from weak democracies to military juntas, which explains the variation of democracy below the 0.6 quantile in Figure 9.

Latin American countries including Chile and Argentina have become more democratic since the 1980's, although several challenges still remain for the region. The experiences of the entire Latin American region in large part demonstrate the first finding in our paper, which was that hybrid regimes are a highly unstable state that leads either to collapse of democracy or to consolidation of it. Over the second part of the 20th century, this has been most obviously the case in Latin America. Once democratic institutions were weakened in Argentina and Chile, the resulting uncertainty quickly gave way to a rapid rise of authoritarianism. Similarly, once the institutions of authoritarianism started to weaken, these same countries' regimes swung

back to the other extreme of established democracies. The cross-country pattern is however best explained by focusing on country-specific effects that our data shows can substantially vary by quantiles of democracy.

East Asia is an example of a region where changing country specific effects are most easily identified. The experiences of China and Taiwan offer one such example (the other being North and South Korea). Around the time of democratization in Latin America, a transition to democracy started in one of the East Asia "miracles", Taiwan, which offered a stunning example of a dual transition both to democracy and to one of the world's best performing economies. In 1988, President Lee Teng-hui initiated a series of reforms to dismantle authoritarian arrangements that years later led to a democratic regime. The transition and achievement of a more democratic government were associated with the levels of democracy predicted by the model beyond the 0.7 quantile in Figure 9.

In contrast, the experience of China has been characterized by a lack of democratic institutions despite unprecedented rates of economic development. In China, country specific institutions including historical legacies, culture and formerly-communist structures of power have changed little despite China's development. In our analysis, China emerges as an outlier and its country specific effects do not change regardless of what quantile of democracy we look at. In fact, China is unique in our sample of Asian countries where the country specific effect remains low as we move from the 0.3 to 0.7 quintile in the conditional distribution of democracy in the world. This result emphasizes what scholars of Chinese politics and history have referred to as the persistence of state institutions and Mao's institutional legacies. For example, the work by Blecher and Shue (1996) has illustrated the ways in which the post-Mao political economy of China has been shaped by his long-lasting legacies such as establishment of "model villages" and state-run enterprises in Northern China. Even slow movements towards more freedom and rights in China have been shaped by the legacies of the past. In-depth historical studies of Chinese social and political movements have shown that protest and resistance against authoritarian government is structured by the persistent legacies of the state (Perry 1980, 2002, 2006; Perry and Xun 1997). In fact, even the concept of political rights in China has been shaped not in line with the classic Western liberal model, but instead by the circumstances and the power of the Chinese state legacies (O'Brien and Li, 2006; Perry and Goldman 2007).

Our data shows that the Chinese experience with slowly-changing country-specific factors is not shared by all Asian countries. Similarly to Taiwan and in contrast to China, South Korea stands out as model a model democracy. In addition to the history of the Park dictatorship in South Korea, one important fact that South Korea and Taiwan share is that they both underwent important land reforms that dramatically redistributed agricultural land, undermining the traditional power of big landowners and opening space for changing structures of political power. While comparable institutional and economic data does not exist for the period before the Second World War, our analysis would predict a similar change in country-specific institutions for Japan, where land reform also played an important role for setting conditions for the stability of a democratic system after the war. However, because of the direct American post-war occupation of

Japan and dramatic changes in its institutions in this period, Japan's country-specific factors in our sample correspond better to the Western European group of countries than to the East Asian.

We conclude this section by cautioning the reader about interpreting our results as overemphasizing the long run forces which the country effects are proxying for. The estimated quantile specific effects are capturing these forces but they also capture many country specific short run events. It is ultimately the heterogeneous nature of idiosyncratic political history, which cannot be captured by the economic and demographic variables included in the structural part of the model. The evidence presented in this section suggests an association between country-specific events and country-specific democratic transitions, which explains the puzzle of the bimodal distribution of democracy. In order to investigate this idea in more detail, we estimate density functions for democracy considering the estimates for the conditional quantiles of democracy. We present the evidence in Figure 10. While the densities for the US and Botswana rank at the top of the democracy distribution, the densities for Cuba and China, not presented in the graph, rank at the bottom. These groups of countries obviously generate a bimodality in the democracy distribution. But a large number of countries, represented here by Greece, Spain, Sierra Leone, Zimbabwe, Chile, Argentina and Taiwan, have density functions with mass mainly at both the bottom and the top of the democracy distribution. Countries with volatile and heterogeneous democratic status also contribute to the explanation behind the puzzle of the bimodal distribution of democracy.

To conclude our brief historical tour, it might perhaps be useful to introduce the notion of democracy as a "foundational institution," akin to what political philosophers refer to as "a foundational good." (Shapiro 1999). While a foundational good is understood as a good that is necessary in order to obtain other goods (as Shapiro argues that from a justice point of view democracy is a foundational good because it allows pursuit of other goods we value), democracy can also be understood as an institution that is necessary for other desirable institutions to flourish. This view is consistent with a power-centered approach to understanding democratic institutions, most notably associated with Madison and Schumpeter. According to Madison's famous line in the Federalist #51, democratic institutions should be structured to allow "ambition to counteract ambition," (Hamilton, Jay and Madison [1788] 1987), a balanced and fair party competition in an electoral democracy can serve to justly allocate political power, protecting individual rights and liberties. Similarly, Joseph Schumpeter developed in 1941 what he called an economic theory of democracy, arguing that just like firms are the major players in a capitalist market, parties should be the major players in a democracy; according to his "minimalist conception" of democracy, the most just political system is that in which government power is allocated according to fair multiparty elections (Schumpeter, 2006). An emphasis on multiparty competition as the basic ingredient of a just pluralist democracy is also made by modern-day theorists such as Rosenblum (2008), according to whom having a strong party system can help channel the distribution of power, ambition and even citizen identity. When considering empirical results reported in this paper, the concept of a minimal procedural democracy as a subordinate foundation good emerges as particularly

attractive since we find that the improvements in these minimal institutional aspects of democracy can produce fundamental change in country's long-run characteristics, opening up opportunities for economic development as well as even better democracy.

The fact that we observe changing country-specific effects as democracy takes hold is evidence that changing democratic institutions are associated with a broader set of political and social changes in a country. Simply put, the best policy for promoting democracy is to focus directly on strengthening procedural and substantive elements of democracy that include those measured by Polity and the Freedom House scores and range from fair and free elections, multiparty competition, representative and accountable government, and a full set of political rights and civil liberties. As regimes improve along these dimensions, our estimation shows that country-specific effects also change. Consequently, both for the proponents of democracy and for economists promoting economic development, a more committed focus on democracy in authoritarian states seems to be the best policy.

7. Conclusion

This paper reports the finding that countries have clustered into two separate and opposite regime types in the second part of the 20th century. The two opposing regime types, authoritarianism and democracy, both emerge as virtually equally prevalent regime types. This indeed may come as a disappointment for proponents of democracy from Herodotus to Vaclav Havel and the modern activists still struggling to bring about revolutions in the world's remaining dictatorships. At first, the persistence of this bimodal distribution of political regimes seems to be explainable by the peculiar inverted-U relationship between income and democracy in the cross-section of all countries 1945-1999. When country-specific effects are left out, hybrid regimes appear to be particularly unstable, possibly representing an unstable equilibrium, where changes in the level of economic development have the largest effect, resulting in democracy when countries perform well and falling back to dictatorships when countries perform poorly.

Our major finding is changing country-specific effects hold the key to political development, understood as the establishment of democratic institutions and all the political rights and civil liberties that come with them. We found that there is virtually zero effect that economic development can have on political institutions when democracy is entirely lacking. The clear policy implication of our finding is that while promoting wealth and education may be a worthwhile cause on its own, higher levels of development should not be expected to lead to better political institutions. Country-specific effects that include the complex milieu of institutions, historical legacies and culture entirely explain why countries have authoritarian governments.

Perhaps ironically, economic development can have an effect on political institutions only when they already start to undergo a movement towards democracy. As democracy develops, our results show that country-specific effects also start to change. This finding runs counter to pessimism that may arise once we realize that

path-dependent historical legacies and entrenched institutions explain virtually all variation in authoritarian states. To the contrary, when democracy starts to take hold, we observe changes in all other institutions, opening up space for economic development to start to bear fruit. The more democracy there is, the more room opens up for income and education to have an effect.

In all, we show that higher income and better education both have strongest effects on political development at the intermediate levels of democratization and almost non-existent in countries at the extremes. This relationship is robust to endogeneity and survives different identification strategies but disappears once we account for country specific effects. Only schooling appears to maintain an effect and only in more democratic countries. In contrast, country-specific idiosyncratic factors are measured at different quantiles and we show that they are not uniformly important across the distribution of democracy. Cross-country heterogeneity overwhelms the potential benefits of economic development at the low quantiles of the democracy distribution, but it diminishes in importance as countries become more democratic. Moreover, the evidence in our sample points to country specific effects capturing both long run effects and more recent political shocks.

The present study conducts a very detailed econometric analysis of the complicated relationship between economic and political development and points to the importance of the relative position of a country within the political spectrum. By focusing on the comparative levels of development we are not addressing the related question as to how countries move between political regimes and how certain regimes become entrenched. A joint analysis of both equilibrium and dynamics of economic development and political institutions remains an important area of future empirical research.

8. **Appendix A: Computational Aspects**

The quantile regression minimization problem can be formulated as a linear program,

$$\min \left\{ \tau \iota_n' r^+ + (1-\tau)\iota_n' r^- | D = Xb + r^+ - r^-, (b, r^+, r^-) \in \mathcal{R}^p \times \mathcal{R}^{2n} \right\} \tag{8.1}$$

where ι is a vector of n ones and $X = [I \dot{:} x]$. The previous problem has a dual formulation,

$$\max \left\{ D'd | X'd = 0, d \in [\tau - 1, \tau]^n \right\}, \tag{8.2}$$

and, for $c = d + 1 - \tau$, may be expressed as,

$$\max \left\{ D'c | X'c = (1-\tau)X'\iota_n, c \in [0,1]^n \right\}. \tag{8.3}$$

Algorithms available in the `quantreg` package of the public domain dialect `R` are based on the previous representation. See Koenker (2004, chapter 6) for a more extensive discussion.

The dual representation provides a way forward in panel data. The primal can be similarly formulated as,

$$\min \left\{ \sum_{j=1}^{J} \omega_j (\tau_j \iota_{NT}' r_j^+ + (1-\tau_j)\iota_{NT}' r_j^-) | \; D = Xb_j + Z\mu + r_j^+ - r_j^-, \;\; j = 1, \ldots, J \right\}, \tag{8.4}$$

or, using a more compact notation,

$$\max \{c'\gamma \mid A'\gamma = (\iota \otimes D)\}, \tag{8.5}$$

where $\gamma = (b_1', \ldots, b_J', a', r_1^+, \ldots, r_J^+, r_1^-, \ldots, r_J^-)'$, and $A = [I_J \otimes X \vdots \iota_J \otimes Z \vdots I_J \otimes I_{NT} \vdots I_J \otimes -I_{NT}]$. Transforming variables similarly as before, we have

$$\max_v \left\{(\iota_J \otimes D)'v \mid B'v = h, v \in [0,1]^{JNT}\right\}, \tag{8.6}$$

with $h = (\omega_1(1-\tau_1)X'\iota_{NT}, \ldots, \omega_J(1-\tau_J)X'\iota_{NT}, \sum_j \omega_j(1-\tau_j)Z'\iota_{NT})'$ and $[0,1]^{JNT}$ denoting the JNT Cartesian product of the closed interval $[0,1]$. Obviously the design matrix is now of large dimension but the problem can be efficiently solved considering the sparse matrix algebra storage used in Koenker (2004).

All programs were written in `Matlab` and `R`. The quantile regression method designed to estimate distributional individual specific effects solves the dual of the problem formulated in 8.4 considering $\mu = \mu_j$. The instrumental variable quantile regression with country effects is a simple extension of the Chernozhukov and Hansen (2005) algorithm that includes individual effects as additional covariates. Results and programs are available upon request.

9. **Appendix B: Data Description**

The data used in the paper spans the period 1945 to 1999, and consists of all countries for which the appropriate data publicly available. The data used has been compiled by Fearon and Laitin (2003), and is provided on their website. We have additionally augmented their dataset by including the Barro-Lee educational achievement variable, available from NBER and the Harvard Center for International Development. This variable is the percentage of the population with secondary education. In a small number of cases linear interpolation was employed to avoid unnecessary gaps in the series when it made sense to replace the missing values with predictions. We measure democracy by the Polity IV score and the Freedom House score, both of which are discussed extensively in the paper.

All historical and geographic data is from Acemoglu et. al. (2001, 2002). Settler mortality was also collected by Acemoglu et. al. (2001). Log population density in 1500 was collected by Acemoglu et. al. (2002). These variables are described in great detail in the above papers.

Additional controls were employed in the sensitivity analysis. Those results were discussed in an earlier version of the paper but are not reproduced in the current version even though they are mentioned. The ethnic fractionalization index is the ethnolinguistic fractionalization index originally reported by the Atlas Naradov Mira and measures the probability that any two individuals in the country speak the same language. The religious fractionalization index is constructed using the same formula estimating the probability of two members of the different groups being of the same religion. Missing values for the index were imputed using additional resources such as the CIA Factbook, Encyclopedia Brittanica, the Library of Congress Studies and

country-specific sources. The oil exporter dummy variable takes the value 1 if more than 33% if a country's exports are based on fossil fuels.

Per capita income is measured in 1985 US dollars and is used in log form, lagged one year. The trade weighted instrument was computed by Acemoglu et. al. (2008) and uses the IMF trade matrix. The world factors instruments are computed from log per capita GDP using Principal Components and employing standard normalizations. The exact procedure is described in the paper.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

——— (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 118, 1231–1294.

——— (2003): "An African Success Story: Botswana," in *In Search of Prosperity: Analytic Narratives on Economic Growth*, ed. by D. Rodrik. Princeton University Press, Princeton.

——— (2005): "The Rise of Europe; Atlantic Trade, Institutional Change and Economic Growth," *American Economic Review*.

Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008): "Income and Democracy," *American Economic Review, forthcoming*.

Algan, Y., and P. Cahuc (2008): "Civic Virtue and Labor Market Institutions," *American Economic Journal: Macroeconomics, forthcoming*.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Proximate Factor Models," *Econometrica*, 70(1), 191–221.

Banac, I. (1988): *The National Question in Yugoslavia: Origins, History, Politics*. Cornell University Press.

Barro, R. J. (1999): "The Determinants of Democracy," *Journal of Political Economy*, 107, S158–S183.

Blanchard, O. (2004): "The Economic Future of Europe," *Journal of Economic Perspectives*, 18(4), 3–26.

Blecher, M., and V. Shue (1996): *Tethered Deer: Government and Economy in a Chinese County*. Stanford University PRess.

Buchinsky, M. (1995): "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–262.

——— (2008): "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 142(1), 379–398.

Cigar, N., P. Williams, and I. Banac (2002): *Indictment at the Hague: The Milosevic Regime and Crimes of the Balkan Wars*. NYU Press.

Collier, P., and A. Hoeffler (2004): "Greed and Grievance in Civil War," *Oxford Economic Papers*.

Colton, T. J. (2008): *Yeltsin: A Life*. Basic Books.

Colton, T. J., and S. Holmes (2007): *The State after Communism: Governance in the New Russia*. Rowman and Littlefield.

Davison, A. C., and D. V. Hinkley (1997): *Bootstrap Methods and Their Applications*. Cambridge University Press.

Ekiert, G., and S. E. Hanson (eds.) (2003): *Capitalism and Democracy in Central and Eastern Europe: Assessing the Legacy of Communist Rule*. Cambridge University Press.

Epstein, D. L., R. H. Bates, J. Gladstone, I. Kristensen, and S. O'Halloran (2006): "Democratic Transitions," *American Journal of Political Science*, 50(3), 551–569.

Fearon, J., and D. Laitin (2003): "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97(1), 75–90.

Fearon, J. D. (2005): "Primary Commodity Exports and Civil War," *Journal of Conflict Resolution*, 49(4), 483–507.

Fish, M. S. (2005): *Democracy Derailed in Russia: The Failure of Open Politics*. Cambridge University Press.

Grzymala-Busse, A. M. (2002): *Redeeming the Communist Past: The Regeneration of Communist Parties in East Central Europe*. Cambridge University Press.

Hahn, J. (1995): "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11(1), 105–121.

Hall, P. A., and D. Soskice (eds.) (2001): *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford University Press, Oxford.

Hamilton, J., J. Madison, and J. Jay (1987): *The Federalist Papers*. Penguin.

Harding, M. C. (2007): "Structural Estimation of High Dimensional Factor Models," Working Paper, MIT.

——— (2008): "Do Global Factors Affect US Inflation Dynamics?," Working Paper, Stanford University.

Horowitz, J. (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327–1352.

Howard, M. M. (2003): *The Weakness of Civil Society in Post-Communist Europe*. Cambridge University Press.

Huntington, S. P. (ed.) (1993): *Transitions from Authoritarian Rule:Southern Europe*. University of Oklahoma Press.

Jones-Luong, P. (ed.) (2002): *Institutional Change and Political Continuity in Post-Soviet Central Asia: Power, Perceptions, and Pacts*. Cambrige University Press.

——— (ed.) (2003): *The Transformation of Central Asia: States and Societies from Soviet Rule to Independence*. Cornell University Press.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Lijphart, A. (1999): *Patterns of democracy : government forms and performance in thirty-six countries*. Yale University Press, New Haven.

Linz, J. J., and A. Stepan (eds.) (1996): *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Johns Hopkins University Press.

Lipset, S. M. (1959): "Some Social Requisites of Democracy: Economic Development and Political Legitimac," *American Political Science Review*, 53, 69–105.

Lipset, S. M. (1963): *Political Man: The Social Bases of Politics*. Anchor Books, New York.

McFaul, M. (2002a): "The Fourth Wave of Democracy and Dictatorship: Noncooperative Transitions in the Postcommunist World," *World Politics*, 54, 212–244.

——— (2002b): *Russia's Unfinished Revolution: Political Change from Gorbachev to Putin*. Cornell University Press.

McFaul, M., and K. Stoner-Weiss (eds.) (2004): *After the Collapse of Communism: Comparative Lessons of Transition*. Cambridge University Press.

Persson, T., G. Roland, and G. Tabellini (2000): "Comparative Politics and Public Finance," *Journal of Political Economy*, 108, 1121–1161.

Persson, T., and G. Tabellini (1997): "Separation of Powers and Political Accountability," *Quarterly Journal of Economics*, 112, 1163–1202.

——— (2000): *Political Economics: Explaining Economic Policy*. Cambridge University Press, Cambridge.

——— (2003): *The Economic Effects of Constitutions: What Do the Data Say?* MIT Press, Cambridge, Mass.

Politkovskaya, A. (2007): *Putin's Russia: Life in a Failing Democracy*. Holt.

Przeworski, A., J. A. Cheibub, M. E. Alvarez, and F. Limongi (2000): *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge University Press, Cambridge.

Przeworski, A., and F. Limongi (1997): "Modernization: Theory and Facts," *World Politics*, 49, 155–183.

Robinson, J. A. (2006): "Economic Development and Democracy," *Annual Reviews of Political Science*, 9, 503–527.

Robinson, J. A., and Q. N. Parsons (2006): "State Formation and Governance in Botswana," *Journal of African Economies*, 15(AERC Sup), 100–140.

Rosenblum, N. L. (2008): *On the Side of the Angels: An Appreciation of Parties and Partisanship*. Princeton University Press.

Schmitter, P. C., L. Whitehead, and G. O'Donnell (eds.) (1986): *Transitions from Authoritarian Rule:Southern Europe*. Johns Hopkins University Press.

Schumpeter, J. (2006): *Capitalism, Socialism and Democracy*. Routledge.

Shapiro, I. (1999): *Democratic Justice*. Yale University Press.

Shevtsova, L. (2007a): *Putin's Russia*. Carnegie Endowment for Peace.

——— (2007b): *Russia - Lost in Transition: The Yeltsin and Putin Legacies*. Carnegie Endowment for Peace.

Stock, J. H., and M. W. Watson (2002): "Has the Business Cycle Changed and Why?," *NBER Macroeconomics Annual*, pp. 191–221.

Variables	Quantiles					Mean
	0.10	0.25	0.50	0.75	0.90	
Polity Measure	Pooled Regressions					
Log GDP per Capita$_{it-1}$	0.006	0.035	0.262	0.185	0.071	0.170
	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)
Log Population	0.010	0.007	0.010	0.002	-0.004	0.021
	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)	(0.003)
Polity Measure	Instrumental Variables Set I					
Log GDP per Capita$_{it-1}$	0.002	0.000	0.369	0.163	0.088	0.222
	(0.006)	(0.008)	(0.667)	(0.056)	(0.005)	(0.007)
Log Population	0.013	0.000	-0.009	-0.016	-0.019	0.004
	(0.002)	(0.002)	(0.247)	(0.005)	(0.004)	(0.003)
Polity Measure	Fixed Effects					
Log GDP per Capita$_{it-1}$	0.001	0.006	-0.003	-0.039	-0.073	0.015
	(0.017)	(0.012)	(0.012)	(0.017)	(0.044)	(0.007)
Log Population	0.073	0.076	0.075	0.080	0.107	0.141
	(0.032)	(0.031)	(0.030)	(0.031)	(0.040)	(0.010)
Freedom House Measure	Pooled Regressions					
Log GDP per Capita$_{it-1}$	0.073	0.119	0.252	0.213	0.145	0.183
	(0.003)	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)
Log Population	0.007	0.018	0.001	-0.015	-0.011	0.011
	(0.003)	(0.005)	(0.002)	(0.002)	(0.001)	(0.003)
Freedom House Measure	Instrumental Variables Set I					
Log GDP per Capita$_{it-1}$	0.069	0.204	0.295	0.217	0.154	0.234
	(0.012)	(0.047)	(0.027)	(0.004)	(0.005)	(0.007)
Log Population	0.004	0.018	-0.016	-0.026	-0.026	0.000
	(0.005)	(0.008)	(0.010)	(0.004)	(0.004)	(0.004)
Freedom House Measure	Fixed Effects					
Log GDP per Capita$_{it-1}$	0.075	0.072	0.058	0.028	0.001	0.060
	(0.035)	(0.035)	(0.035)	(0.037)	(0.038)	(0.012)
Log Population	0.072	0.076	0.078	0.075	0.072	0.156
	(0.037)	(0.037)	(0.038)	(0.039)	(0.039)	(0.013)

TABLE 1. Regression results for the full sample of countries. This sample includes 156 countries and 6220 observations. The instrumental variable set includes log of mountainous terrain, geographic latitude, and log of air distance to nearest port. Standard errors in parenthesis.

Variables	Quantiles					Mean
	0.10	0.25	0.50	0.75	0.90	
Polity Measure	Pooled Regressions					
Log GDP per Capita$_{it-1}$	-0.005	0.064	0.158	0.108	0.054	0.092
	(0.003)	(0.007)	(0.008)	(0.007)	(0.003)	(0.008)
Log Population	0.008	0.002	-0.017	-0.007	-0.005	-0.003
	(0.000)	(0.004)	(0.004)	(0.002)	(0.001)	(0.003)
Schooling	0.015	0.044	0.049	0.013	0.001	0.043
	(0.001)	(0.006)	(0.003)	(0.002)	(0.001)	(0.003)
Polity Measure	Instrumental Variables Set I					
Log GDP per Capita$_{it-1}$	-0.019	0.007	0.209	0.149	0.134	0.043
	(0.087)	(0.032)	(0.025)	(0.023)	(0.014)	(0.032)
Log Population	-0.008	-0.032	-0.029	-0.021	-0.022	-0.028
	(0.022)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)
Schooling	0.034	0.084	0.031	0.006	-0.013	0.060
	(0.021)	(0.008)	(0.009)	(0.007)	(0.004)	(0.009)
Polity Measure	Fixed Effects					
Log GDP per Capita$_{it-1}$	-0.027	-0.019	-0.017	-0.074	-0.181	-0.048
	(0.045)	(0.032)	(0.029)	(0.038)	(0.064)	(0.013)
Log Population	0.091	0.093	0.092	0.091	0.123	0.037
	(0.063)	(0.064)	(0.064)	(0.065)	(0.069)	(0.019)
Schooling	0.041	0.034	0.028	0.033	0.037	0.074
	(0.022)	(0.020)	(0.021)	(0.023)	(0.028)	(0.006)
Freedom House Measure	Pooled Regressions					
Log GDP per Capita$_{it-1}$	0.030	0.068	0.167	0.164	0.116	0.111
	(0.008)	(0.011)	(0.008)	(0.006)	(0.004)	(0.008)
Log Population	-0.020	-0.009	-0.014	-0.018	-0.013	-0.012
	(0.003)	(0.005)	(0.003)	(0.002)	(0.002)	(0.003)
Schooling	0.037	0.069	0.035	0.009	0.008	0.041
	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)
Freedom House Measure	Instrumental Variables Set I					
Log GDP per Capita$_{it-1}$	-0.007	0.244	0.267	0.212	0.173	0.189
	(0.023)	(0.033)	(0.021)	(0.156)	(0.015)	(0.024)
Log Population	-0.033	-0.029	-0.029	-0.032	-0.034	-0.029
	(0.006)	(0.005)	(0.004)	(0.008)	(0.004)	(0.004)
Schooling	0.054	0.033	0.005	-0.010	-0.002	0.021
	(0.008)	(0.011)	(0.007)	(0.044)	(0.004)	(0.007)
Freedom House Measure	Fixed Effects					
Log GDP per Capita$_{it-1}$	0.063	0.059	0.047	0.005	-0.041	0.011
	(0.044)	(0.042)	(0.041)	(0.041)	(0.044)	(0.017)
Log Population	-0.005	-0.009	-0.004	-0.006	-0.009	-0.036
	(0.064)	(0.065)	(0.065)	(0.065)	(0.065)	(0.029)
Schooling	0.035	0.033	0.028	0.034	0.049	0.062
	(0.019)	(0.019)	(0.019)	(0.022)	(0.025)	(0.007)

TABLE 2. Regression results of a model with education for the full sample of countries. This sample includes 110 countries and 4229 observations. The instrumental variable set includes log of mountainous terrain, geographic latitude, and log of air distance to nearest port. Standard errors in parenthesis.

Variables	Quantiles				
	0.10	0.25	0.50	0.75	0.90
Polity Measure	Instrumental Variables Set I				
Log GDP per Capita$_{it-1}$	0.002	0.000	0.369	0.163	0.088
	(0.006)	(0.008)	(0.667)	(0.056)	(0.005)
Log Population	0.013	0.000	-0.009	-0.016	-0.019
	(0.002)	(0.002)	(0.247)	(0.005)	(0.004)
Polity Measure	Instrumental Variable Set II				
Log GDP per Capita$_{it-1}$	0.158	0.126	0.261	0.302	0.000
	(0.014)	(0.010)	(0.013)	(0.026)	(0.058)
Log Population	0.001	0.000	-0.009	-0.010	0.000
	(0.004)	(0.003)	(0.004)	(0.006)	(0.004)
Polity Measure	Instrumental Variable Set III				
Log GDP per Capita$_{it-1}$	0.013	0.000	0.178	0.054	0.000
	(0.009)	(0.015)	(0.033)	(0.017)	(0.011)
Log Population	0.013	0.000	0.049	0.008	0.000
	(0.003)	(0.002)	(0.019)	(0.003)	(0.002)
Freedom House Measure	Instrumental Variables Set I				
Log GDP per Capita$_{it-1}$	0.069	0.204	0.295	0.217	0.154
	(0.012)	(0.047)	(0.027)	(0.004)	(0.005)
Log Population	0.004	0.018	-0.016	-0.026	-0.026
	(0.005)	(0.008)	(0.010)	(0.004)	(0.004)
Freedom House Measure	Instrumental Variable Set II				
Log GDP per Capita$_{it-1}$	0.141	0.107	0.136	0.170	0.000
	(0.010)	(0.009)	(0.063)	(0.020)	(0.608)
Log Population	-0.004	0.003	-0.007	-0.017	0.000
	(0.005)	(0.005)	(0.009)	(0.004)	(0.079)
Freedom House Measure	Instrumental Variable Set III				
Log GDP per Capita$_{it-1}$	0.086	0.330	0.292	0.390	0.000
	(0.021)	(0.015)	(0.012)	(0.030)	(0.074)
Log Population	0.014	0.013	-0.007	-0.010	0.000
	(0.007)	(0.004)	(0.005)	(0.005)	(0.011)

TABLE 3. Sensitivity analysis for the instrumental variables. The first set of instruments includes log of mountainous terrain, geographic latitude, log of air distance to nearest port; the second set of instruments includes trade-weighted world income as in Acemoglu et. al. (2008); the third set includes the five instrumental variables presented in the appendix. Standard errors in parenthesis.

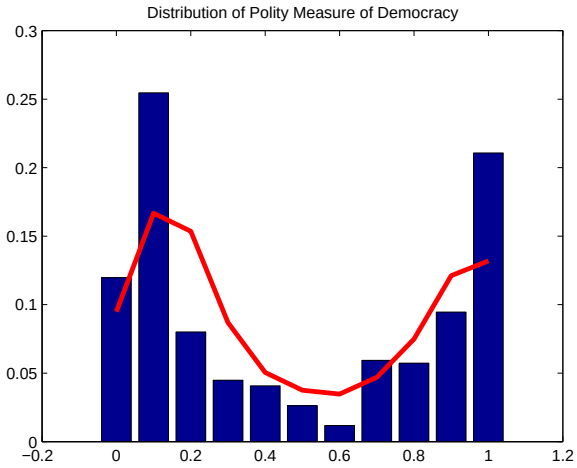

FIGURE 1. Distribution of the normalized Polity IV score for the period 1945-1999. Scale 1 represents the highest level of democracy. The continuous line denotes a kernel density estimate.

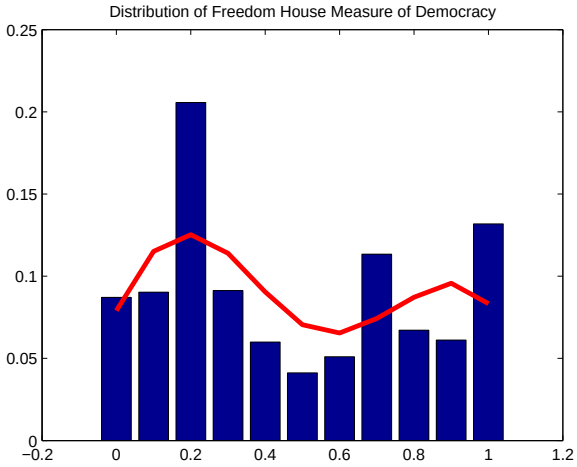

FIGURE 2. Distribution of the normalized Freedom House (reversed) rating for the period 1972-1999. Scale 1 represents the highest level of democracy. The continuous line denotes a kernel density estimate.

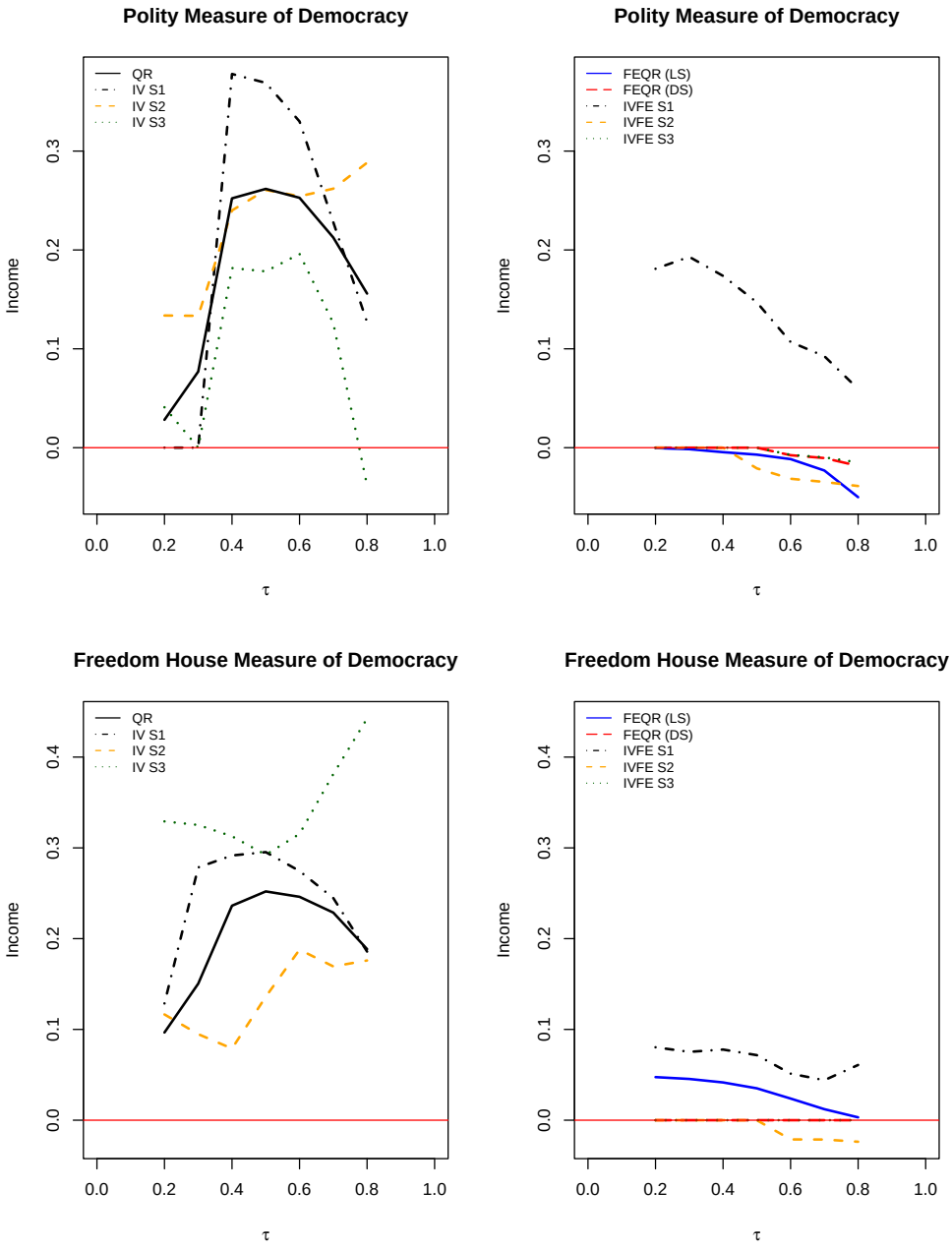

FIGURE 3. Quantile regression estimates of the effect of income on democracy considering the full sample of countries. The panels show point estimates obtained by using quantile regression for the pooled data (QR), instrumental variable with and without fixed effects (IVFE, IV), and quantile regression with fixed effects assuming that the individual effects are either location shifts (LS) or distributional shifts (DS). The instrument sets (S1-S3) are described in the paper.

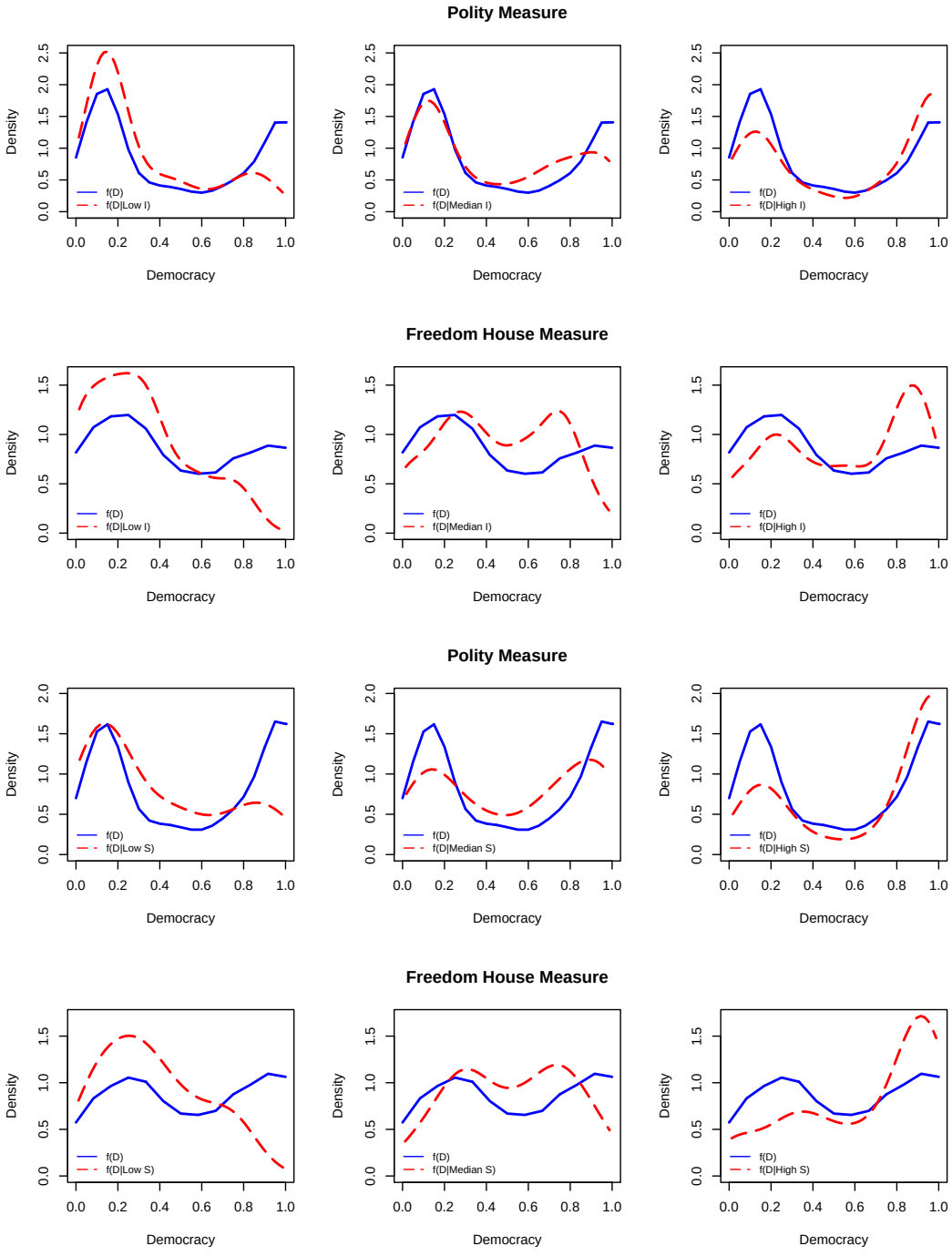

FIGURE 4. Conditional and Unconditional distributions of the measures of democracy. The continuous lines denote kernel density estimates of the unconditional distribution and the dashed lines show kernel density estimates for the conditional distribution. The conditioning variables are Income (I) and Schooling (S).

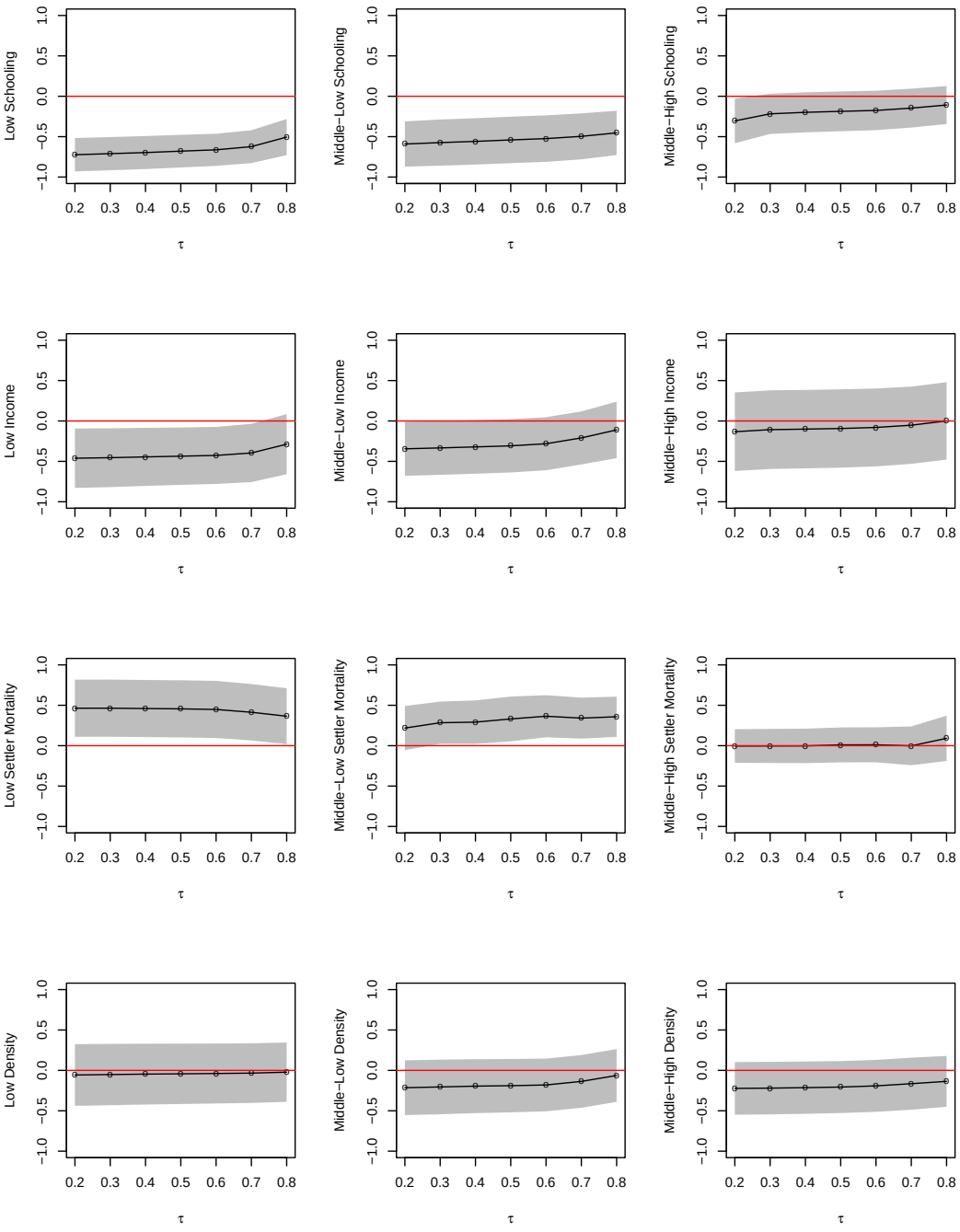

FIGURE 5. Group effect estimates in a model for the Polity IV measure of democracy. The model also includes log of GDP at t-1, log of population, and country specific location shifts. The continuous line with dots represents the quantile regression group effects estimates and the shaded grey area represents a .95 pointwise confidence interval. Groups of countries are defined by comparing the mean level of the country's variable to the quartiles of the variable's distribution (e.g., low schooling means that the average schooling is below the $25th$ percentile). The composition of the groups is described in the appendix.

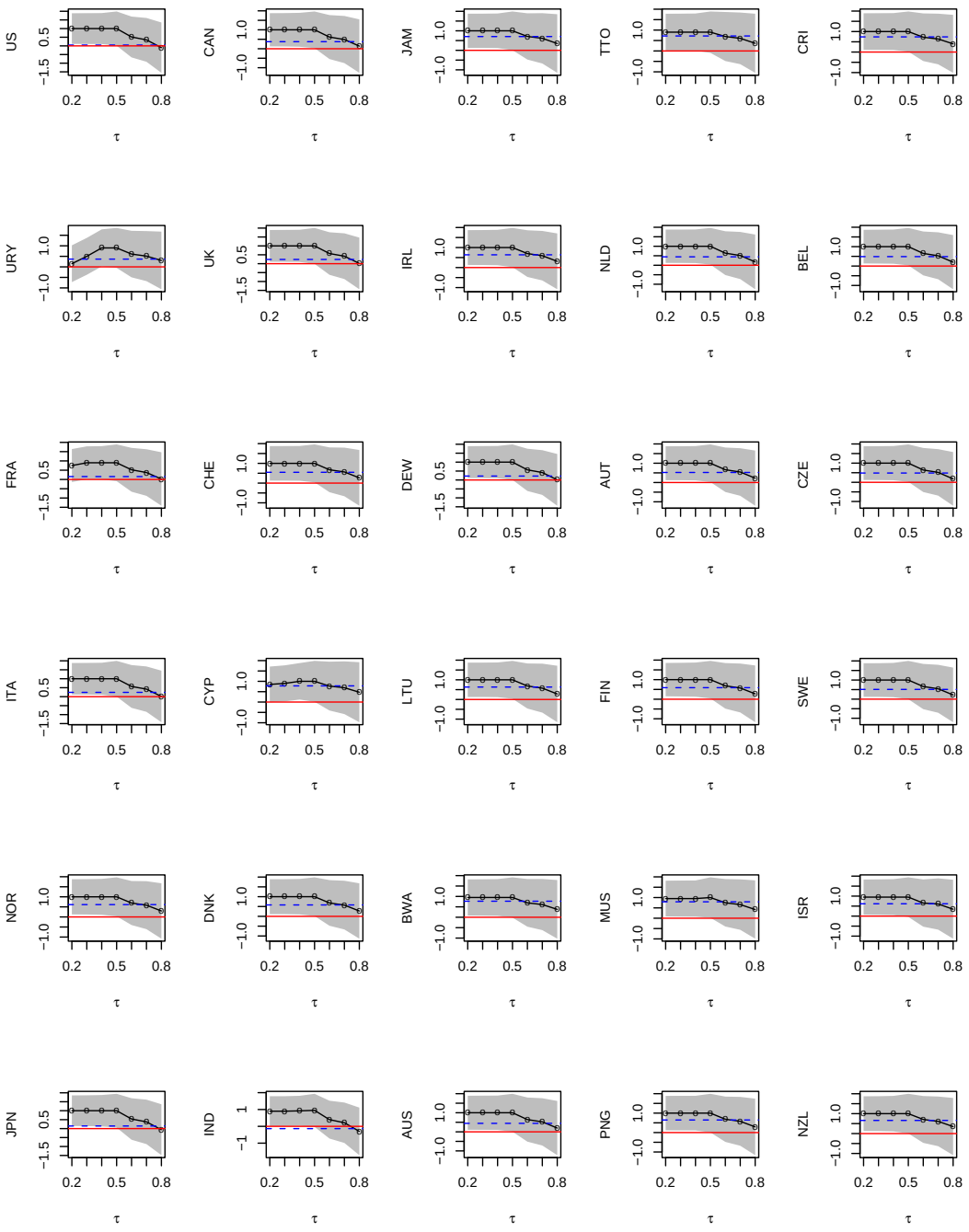

FIGURE 6. Country effect estimates in a model for the Polity IV measure of democracy. The figure presents results for countries with at least one individual effect significant at 1 percent, in a model that includes log of GDP at t-1 and log of population. The continuous line with dots represents the quantile regression country effects estimates and the dashed blue line the mean fixed effect. The shaded grey area represents a pointwise confidence interval.

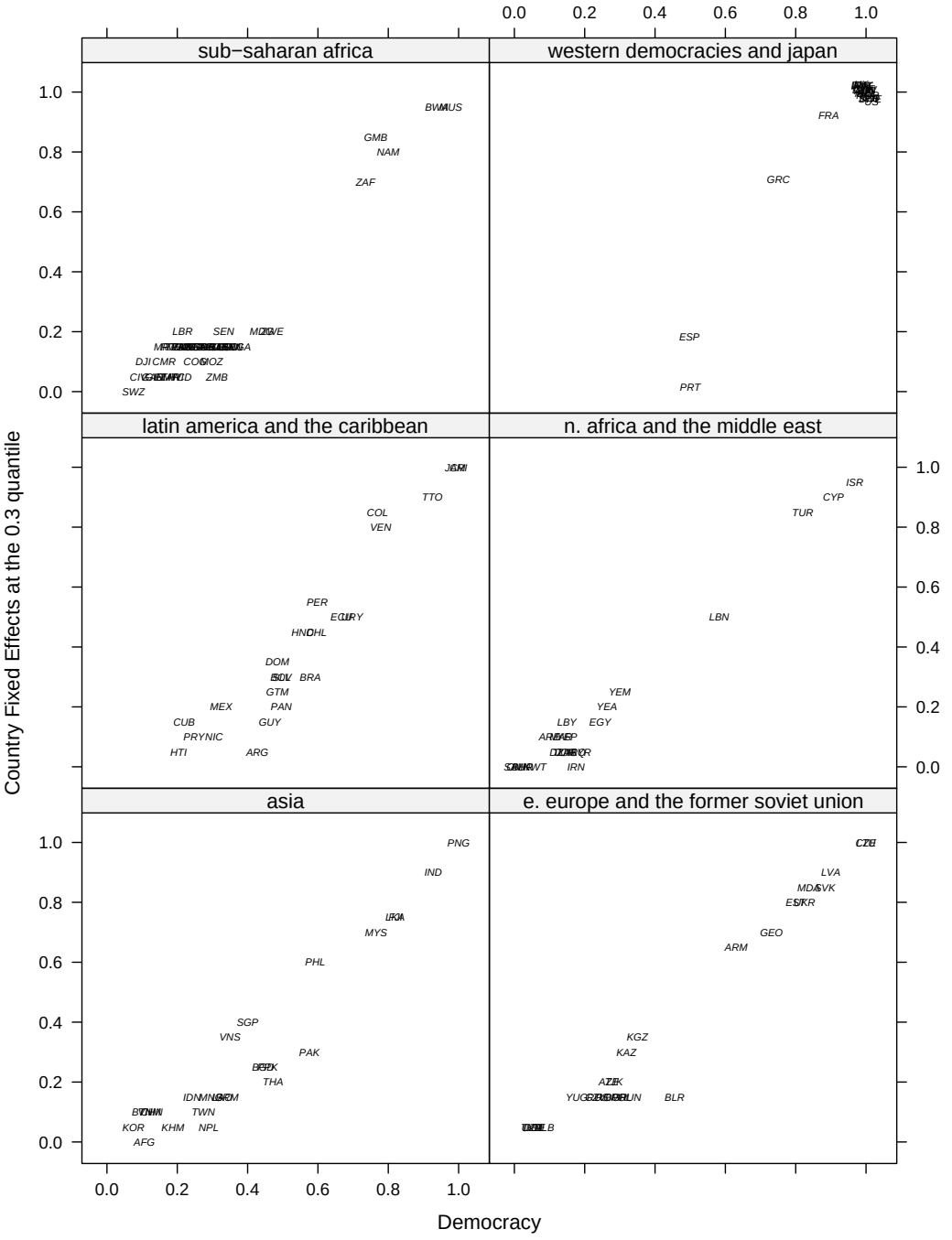

FIGURE 7. Scatterplot of country effects and democracy levels. The country effects were estimated at the 0.3 quantile of the conditional distribution of democracy, in a linear model that includes log of DGP at $t-1$, log of population, and schooling. The dependent variable democracy is constructed based on the Polity measure. See the appendix for data definitions.

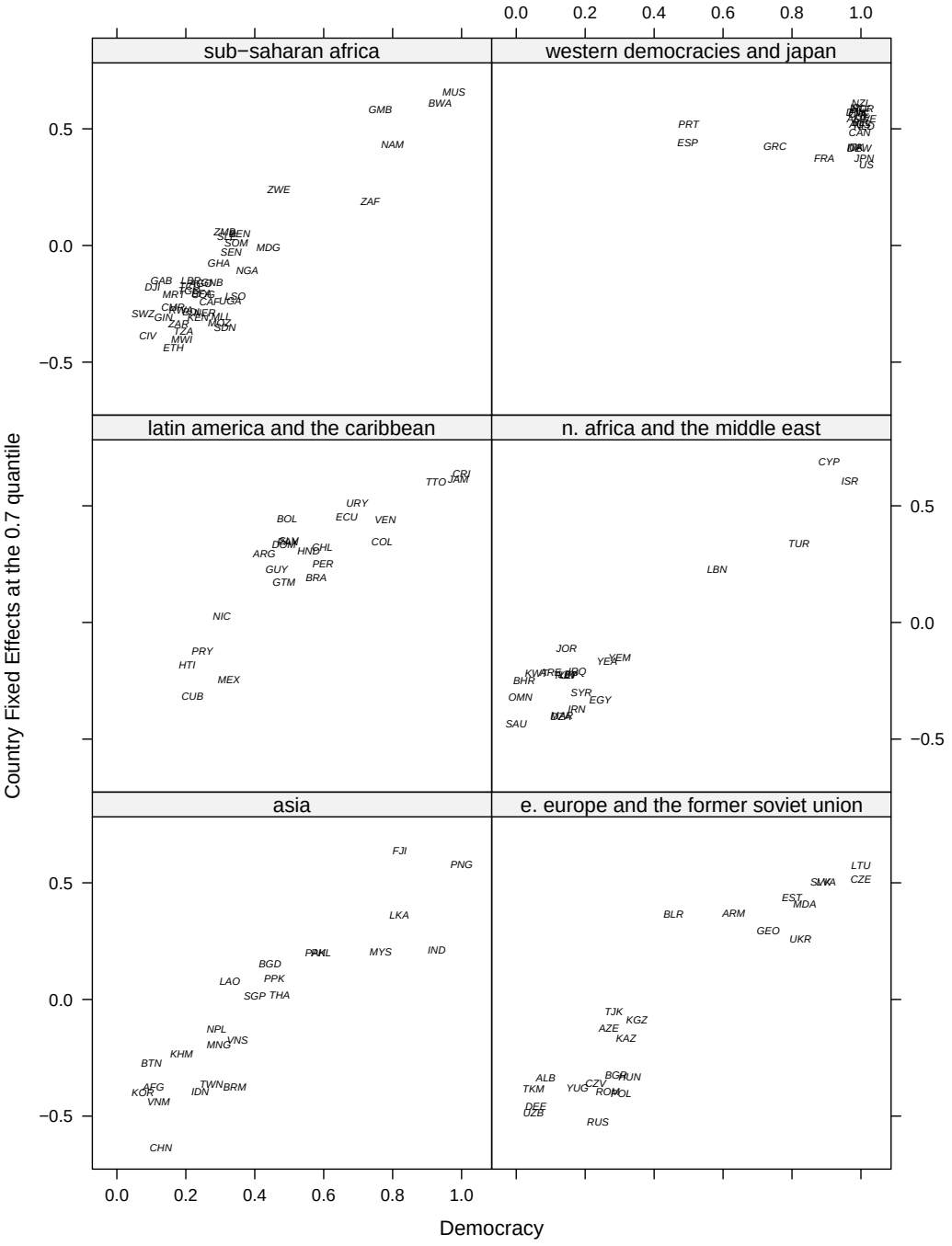

FIGURE 8. Scatterplot of country effects and democracy levels. The country effects were estimated at the 0.7 quantile of the conditional distribution of democracy, in a linear model that includes log of DGP at $t-1$, log of population, and schooling. The dependent variable democracy is constructed based on the Polity measure. See the appendix for data definitions.

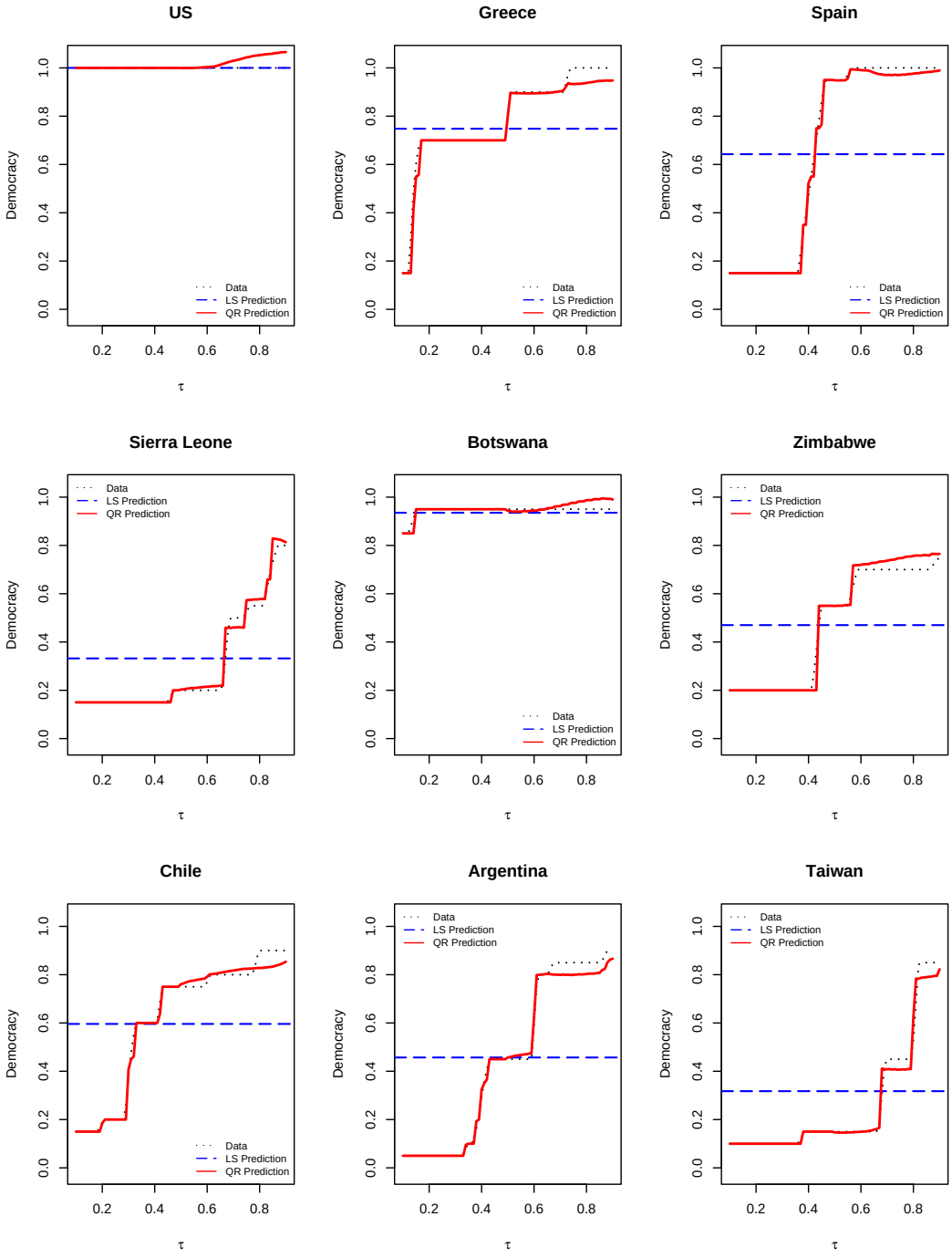

FIGURE 9. Empirical conditional quantiles and quantiles of democracy distribution for some countries representing different regions of the world. For each country, three lines are presented indicating the quantiles of democracy D (dotted line), the conditional mean $\hat{E}(D|I,x,\mu)$ (dashed line), and the conditional quantiles $\hat{Q}_D(\tau|I,x,\mu)$ (continuous line). The regression models use the country-specific mean of log of GDP and log of population as well as country specific effects.

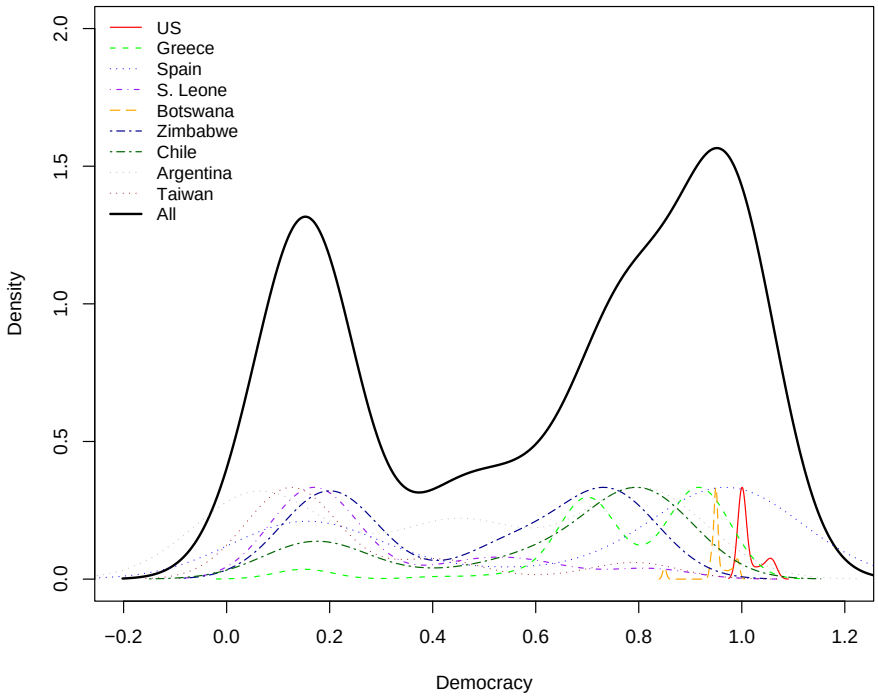

FIGURE 10. Explaining the puzzle. The figure shows a density function for the conditional quantiles of democracy in nine countries representing different regions of the world (black, continuous line). The figure also shows (rescaled) country-specific density functions based on conditional quantiles $\hat{Q}_D(\tau|I, x, \mu)$. The quantile regression model uses the country-specific mean of log of GDP and log of population as well as country specific effects.

The Political Economy of Heterogeneous Development: Quantile Effects of Income and Education*

Marcus Alexander†, Matthew Harding‡ and Carlos Lamarche§

August 2008

Online Appendix

Abstract

Does development lead to the establishment of more democratic institutions? The key to the puzzle, we argue, is the previously unrecognized fact that based on quantitative regime scores, countries over the past 50 years have clustered into two separate, very distinct, yet equally-common stages of political development — authoritarian states with low levels of freedom on one side and democracies with liberal institutions on the other side of a bimodal distribution of political regimes. We develop a new empirical strategy — exploiting exogenous world economic factors and introducing new panel data estimators — that allows for the first time to estimate the effects of development as well as changing unobserved country effects in driving democracy at these different stages of political development. We find that income and education have the least effect on democracy when authoritarian regimes are consolidated and that only changing country effects, possibly accounting for institutional legacies, can lead to political development. Ironically, it is in highly democratic and wealthiest of nations that income and education start to play a role; however greater wealth and better educated citizenry can both help and hurt democracy depending again on what the country's institutional legacies are. Far from accepting the notion that much of the developing world is cursed by unchanging and poor long-run institutions, policy-makers should take note that with democratization we also see changing country-specific factors that in turn condition the difference income and education make for democracy.

JEL: C13, C23, P16, O10
Keywords: democracy, economic development, quantile regression

*We are grateful to Tim Hyde for excellent research assistance. We are grateful to Robert Bates, Victor Chernozhukov, Jorge Dominguez, Dawn Brancati, Margarita Estavez-Abe, Kevin Grier, Yoshiko Herrera, Nahomi Ichino, Torben Iversen, Gary King, Roderick MacFarquhar, Thomas Remington, James Robinson and seminar participants at Harvard and Stanford for useful comments, and to Daron Acemoglu, James Robinson, Simon Johnson and Pierre Yared for sharing their historical data.

†Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138; Email: malexand@fas.harvard.edu

‡Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: mch@stanford.edu

§Department of Economics, University of Oklahoma, 729 Elm Avenue, Norman, OK 73019; Phone: (405) 325-5857; Email: lamarche@ou.edu

Country		Obs	From	To	Country		Obs	From	To
Afghanistan	AFG	55	1945	1999	Dominican Rep.	DOM	55	1945	1999
Albania	ALB	55	1945	1999	Ecuador	ECU	55	1945	1999
Algeria	DZA	38	1962	1999	Egypt	EGY	55	1945	1999
Angola	AGO	25	1975	1999	El Salvador	SLV	55	1945	1999
Argentina	ARG	55	1945	1999	Eritrea	ERI	7	1993	1999
Armenia	ARM	9	1991	1999	Estonia	EST	9	1991	1999
Australia	AUS	55	1945	1999	Ethiopia	ETH	55	1945	1999
Austria	AUT	55	1945	1999	Fiji	FJI	30	1970	1999
Azerbaijan	AZE	9	1991	1999	Finland	FIN	55	1945	1999
Bahrain	BHR	29	1971	1999	France	FRA	55	1945	1999
Bangladesh	BGD	28	1972	1999	Gabon	GAB	40	1960	1999
Belarus	BLR	9	1991	1999	Gambia	GMB	35	1965	1999
Belgium	BEL	55	1945	1999	Georgia	GEO	9	1991	1999
Benin	BEN	40	1960	1999	Germany, Dem. Rep.	DEE	42	1949	1990
Bhutan	BTN	29	1971	1999	Germany, Fed. Rep.	DEW	55	1945	1999
Bolivia	BOL	55	1945	1999	Ghana	GHA	43	1957	1999
Bosnia	BIH	8	1992	1999	Greece	GRC	55	1945	1999
Botswana	BWA	34	1966	1999	Guatemala	GTM	55	1945	1999
Brazil	BRA	55	1945	1999	Guinea	GIN	42	1958	1999
Bulgaria	BGR	55	1945	1999	Guinea Bissau	GNB	26	1974	1999
Burkina Faso	BFA	40	1960	1999	Guyana	GUY	34	1966	1999
Burma	BRM	52	1948	1999	Haiti	HTI	55	1945	1999
Burundi	BDI	38	1962	1999	Honduras	HND	55	1945	1999
Cambodia	KHM	47	1953	1999	Hungary	HUN	55	1945	1999
Cameroon	CMR	40	1960	1999	India	IND	53	1947	1999
Canada	CAN	55	1945	1999	Indonesia	IDN	51	1949	1999
Central African Rep.	CAF	40	1960	1999	Iran	IRN	55	1945	1999
Chad	TCD	40	1960	1999	Iraq	IRQ	55	1945	1999
Chile	CHL	55	1945	1999	Ireland	IRL	55	1945	1999
China	CHN	55	1945	1999	Israel	ISR	52	1948	1999
Colombia	COL	55	1945	1999	Italy	ITA	55	1945	1999
Congo	ZAR	40	1960	1999	Ivory Coast	CIV	40	1960	1999
Costa Rica	CRI	55	1945	1999	Jamaica	JAM	38	1962	1999
Croatia	HRV	8	1992	1999	Japan	JPN	55	1945	1999
Cuba	CUB	55	1945	1999	Jordan	JOR	54	1946	1999
Cyprus	CYP	40	1960	1999	Kazakhstan	KAZ	9	1991	1999
Czechoslovakia	CZV	48	1945	1992	Kenya	KEN	37	1963	1999
Czech, Rep	CZE	7	1993	1999	S. Korea	PPK	51	1949	1999
Congo, Dem. Rep.	COG	40	1960	1999	Kuwait	KWT	39	1961	1999
Denmark	DNK	55	1945	1999	Kyrgyzstan	KGZ	9	1991	1999
Djibouti	DJI	23	1977	1999	Laos	LAO	47	1953	1999

TABLE 1. Sample of Countries.

Country		Obs	From	To	Country		Obs	From	To
Latvia	LVA	9	1991	1999	Sierra Leone	SLE	39	1961	1999
Lebanon	LBN	54	1946	1999	Singapore	SGP	35	1965	1999
Lesotho	LSO	34	1966	1999	Slovakia	SVK	7	1993	1999
Liberia	LBR	55	1945	1999	Slovenia	SVN	8	1992	1999
Libya	LBY	49	1951	1999	Somalia	SOM	40	1960	1999
Lithuania	LTU	9	1991	1999	South Africa	ZAF	55	1945	1999
Macedonia	MKD	7	1993	1999	Spain	ESP	55	1945	1999
Madagascar	MDG	40	1960	1999	Sri Lanka	LKA	52	1948	1999
Malawi	MWI	36	1964	1999	Sudan	SDN	44	1956	1999
Malaysia	MYS	43	1957	1999	Swaziland	SWZ	32	1968	1999
Mali	MLI	40	1960	1999	Sweden	SWE	55	1945	1999
Mauritania	MRT	40	1960	1999	Switzerland	CHE	55	1945	1999
Mauritius	MUS	32	1968	1999	Syria	SYR	54	1946	1999
Mexico	MEX	55	1945	1999	Taiwan	TWN	51	1949	1999
Moldova	MDA	9	1991	1999	Tajikistan	TJK	9	1991	1999
Mongolia	MNG	55	1945	1999	Tanzania	TZA	39	1961	1999
Morocco	MAR	44	1956	1999	Thailand	THA	55	1945	1999
Mozambique	MOZ	25	1975	1999	Togo	TGO	40	1960	1999
N. Korea	KOR	52	1948	1999	Trinidad & Tobago	TTO	38	1962	1999
Namibia	NAM	10	1990	1999	Tunisia	TUN	44	1956	1999
Nepal	NPL	55	1945	1999	Turkey	TUR	55	1945	1999
Netherlands	NLD	55	1945	1999	Turkmenistan	TKM	9	1991	1999
New Zealand	NZL	55	1945	1999	United Arab Emirates	ARE	29	1971	1999
Nicaragua	NIC	55	1945	1999	Uganda	UGA	38	1962	1999
Niger	NER	40	1960	1999	UK	GBR	55	1945	1999
Nigeria	NGA	40	1960	1999	Ukraine	UKR	9	1991	1999
Norway	NOR	55	1945	1999	Uruguay	URY	55	1945	1999
Oman	OMN	29	1971	1999	USA	USA	55	1945	1999
Pakistan	PAK	53	1947	1999	Uzbekistan	UZB	9	1991	1999
Panama	PAN	55	1945	1999	Venezuela	VEN	55	1945	1999
Papua Nueva Guinea	PNG	25	1975	1999	Vietnam	VNM	46	1954	1999
Paraguay	PRY	55	1945	1999	Vietnam, S.	VNS	22	1954	1975
Peru	PER	55	1945	1999	Yemen	YEM	10	1990	1999
Philippines	PHL	54	1946	1999	Yemen, Arab. Rep.	YEA	46	1945	1990
Poland	POL	55	1945	1999	Yemen, Peop. Rep.	YEP	24	1967	1990
Portugal	PRT	55	1945	1999	Yugoslavia	YUG	55	1945	1999
Romania	ROM	55	1945	1999	Zambia	ZMB	36	1964	1999
Russia	RUS	55	1945	1999	Zimbabwe	ZWE	35	1965	1999
Rwanda	RWA	38	1962	1999					
Saudi Arabia	SAU	55	1945	1999					
Senegal	SEN	40	1960	1999					

TABLE 2. Sample of Countries (Cont.).

Variables	Sample of Countries					
	All countries	West & Japan	E. Europe & Soviet Union	Latin America	Sub-saharan Africa	Asia
Polity Measure of Democracy	0.476 (0.376)	0.931 (0.217)	0.325 (0.314)	0.537 (0.335)	0.313 (0.284)	0.409 (0.332)
Freedom House Measure of Democracy	0.465 (0.333)	0.955 (0.107)	0.416 (0.309)	0.587 (0.264)	0.287 (0.236)	0.374 (0.279)
Log of GDP per capita	7.660 (1.054)	8.939 (0.591)	7.684 (0.884)	7.780 (0.595)	6.772 (0.630)	7.129 (0.803)
Log of Population	9.049 (1.457)	9.595 (1.213)	9.431 (1.223)	8.643 (1.218)	8.500 (1.211)	9.923 (1.761)
Schooling	4.631 (2.911)	7.915 (2.085)	7.978 (1.508)	4.095 (1.672)	2.057 (1.481)	3.626 (2.282)
Log of mountainous terrain	2.177 (1.404)	1.991 (1.435)	1.931 (1.295)	2.645 (1.204)	1.558 (1.435)	2.818 (1.205)
Absolute geographic latitude	0.291 (0.189)	0.533 (0.107)	0.536 (0.062)	0.175 (0.097)	0.131 (0.084)	0.244 (0.140)
Log of air distance from capital to nearest port	7.958 (0.969)	6.628 (1.156)	7.251 (0.658)	8.356 (0.437)	8.695 (0.240)	8.223 (0.494)
Settler Mortality	211.1 (395.1)	8.938 (5.265)	-	88.159 (32.048)	460.2 (588.5)	90.40 (51.81)
Population Density in 1500	6.998 (9.712)	15.044 (13.363)	11.797 (10.434)	1.282 (0.894)	3.378 (5.169)	10.032 (10.657)
Trade-Weighted World Income	10.761 (8.049)	11.377 (4.784)	8.466 (2.169)	10.501 (6.143)	9.396 (5.166)	13.455 (16.649)
Instrumental Variable 1	-1.923 (4.745)	-3.236 (3.251)	-1.716 (5.317)	-1.610 (4.588)	-1.135 (5.482)	-1.955 (4.658)
Instrumental Variable 2	-5.072 (6.232)	-4.540 (6.019)	-5.230 (5.545)	-4.794 (6.049)	-6.022 (6.487)	-4.281 (6.747)
Instrumental Variable 3	10.072 (6.050)	9.789 (4.802)	11.805 (5.584)	10.156 (4.857)	9.425 (8.205)	9.946 (5.696)
Instrumental Variable 4	-3.010 (7.242)	-3.356 (6.667)	-0.509 (7.951)	-3.329 (6.788)	-3.262 (7.757)	-3.271 (7.056)
Instrumental Variable 5	34.308 (13.051)	30.767 (13.751)	35.223 (13.831)	32.036 (13.791)	38.437 (10.315)	33.985 (13.033)

TABLE 3. Descriptive Statistics.

Variables	Quantiles					Mean
	0.10	0.25	0.50	0.75	0.90	
Polity Measure	Pooled Regressions					
Log GDP per Capita$_{it-1}$	0.092	0.246	0.278	0.158	0.059	0.227
	(0.011)	(0.012)	(0.010)	(0.019)	(0.005)	(0.014)
Log Population	-0.002	0.007	-0.005	-0.006	-0.004	0.002
	(0.003)	(0.003)	(0.004)	(0.003)	(0.001)	(0.016)
Polity Measure	Instrumental Variables					
Log GDP per Capita$_{it-1}$	0.000	0.000	0.147	0.290	0.000	0.145
	(0.071)	(0.047)	(0.059)	(0.122)	(0.393)	(0.049)
Log Population	0.000	0.000	-0.003	-0.008	0.000	0.002
	(0.007)	(0.009)	(0.010)	(0.015)	(0.006)	(0.007)
Polity Measure	Fixed Effects					
Log GDP per Capita$_{it-1}$	0.020	0.016	0.012	-0.012	-0.058	-0.023
	(0.030)	(0.029)	(0.030)	(0.029)	(0.035)	(0.032)
Log Population	0.086	0.089	0.088	0.088	0.090	0.074
	(0.040)	(0.040)	(0.039)	(0.039)	(0.038)	(0.061)
Freedom House Measure	Pooled Regressions					
Log GDP per Capita$_{it-1}$	0.193	0.302	0.269	0.199	0.091	0.234
	(0.012)	(0.012)	(0.010)	(0.011)	(0.010)	(0.012)
Log Population	0.002	-0.001	-0.012	-0.019	-0.009	-0.008
	(0.003)	(0.007)	(0.005)	(0.003)	(0.002)	(0.011)
Freedom House Measure	Instrumental Variables					
Log GDP per Capita$_{it-1}$	0.000	0.135	0.201	0.000	0.000	0.146
	(0.110)	(0.044)	(0.087)	(0.445)	(0.298)	(0.046)
Log Population	0.000	0.001	-0.017	0.000	0.000	-0.009
	(0.012)	(0.011)	(0.006)	(0.029)	(0.029)	(0.005)
Freedom House Measure	Fixed Effects					
Log GDP per Capita$_{it-1}$	0.103	0.092	0.070	0.036	0.005	0.016
	(0.039)	(0.034)	(0.033)	(0.034)	(0.036)	(0.030)
Log Population	-0.025	-0.013	-0.010	-0.009	-0.015	-0.103
	(0.032)	(0.031)	(0.031)	(0.030)	(0.030)	(0.059)

TABLE 4. Sensitivity analysis using the 5-year data in Acemoglu et. al. (2008). The instrument for income is trade weighted world income. Year effects are included in all the regressions. Standard errors in parenthesis.

Country	Educ	Income	Settler Mortality	Pop Density	Country	Educ	Income	Settler Mortality	Pop Density
AFG	Low	Low	M-High	M-Low	DOM	M-Low	M-Low	M-High	M-Low
ALB	-	Low	-	M-High	ECU	M-High	M-Low	M-Low	M-Low
DZA	Low	M-High	M-High	Low	EGY	M-Low	M-Low	Low	M-High
AGO	-	Low	High	M-Low	SLV	M-Low	M-Low	M-Low	M-Low
ARG	M-High	High	Low	Low	ERI	-	-	-	M-Low
ARM	-	M-High	-	M-High	EST	High	M-High	-	Low
AUS	High	High	Low	Low	ETH	Low	Low	Low	M-Low
AUT	High	High	-	High	FJI	M-High	M-High	Low	
AZE	-	M-Low	-	M-High	FIN	High	High	-	Low
BHR	M-Low	High	-	M-Low	FRA	M-High	High	Low	High
BGD	Low	M-Low	M-Low	High	GAB	-	M-High	High	M-Low
BLR	-	M-High	-	Low	GMB	Low	Low	High	M-High
BEL	High	High	-	High	GEO	-	M-Low	-	M-High
BEN	Low	M-Low	High	M-High	DEE	High	High	-	High
BTN	-	Low	-	-	DEW	High	High	-	High
BOL	M-High	M-Low	M-Low	Low	GHA	M-Low	Low	High	M-High
BIH	-	-	-	High	GRC	M-High	M-High	-	M-High
BWA	M-Low	M-Low	-	Low	GTM	Low	M-Low	M-Low	M-Low
BRA	M-Low	M-High	M-Low	Low	GIN	-	Low	High	M-High
BGR	High	M-Low	-	M-High	GNB	-	Low	M-High	M-High
BFA	-	Low	High	M-High	GUY	M-High	M-Low	Low	Low
BRM	Low	Low	Low	M-High	HTI	Low	Low	M-High	M-Low
BDI	Low	Low	High	High	HND	M-Low	M-Low	M-Low	M-Low
KHM	-	Low	-	High	HUN	High	M-High	-	High
CMR	Low	M-Low	High	M-Low	IND	M-Low	Low	Low	High
CAN	High	High	Low	Low	IDN	M-Low	M-Low	M-High	M-High
CAF	Low	Low	High	M-Low	IRN	Low	M-High	-	M-Low
TCD	-	Low	High	Low	IRQ	Low	M-High	-	M-Low
CHL	M-High	M-High	Low	Low	IRL	High	High	-	High
CHN	M-High	Low	M-High	High	ISR	High	High	-	M-High
COL	M-Low	M-High	M-Low	Low	ITA	M-High	High	-	High
ZAR	M-Low	M-Low	M-High	M-Low	CIV	-	M-Low	High	M-High
CRI	M-High	M-High	M-Low	M-Low	JAM	M-Low	M-High	M-High	M-High
HRV	M-High	-	-	High	JPN	High	High	-	High
CUB	M-High	M-High	-	Low	JOR	M-Low	M-Low	-	M-High
CYP	M-High	High	-	-	KAZ	High	M-High	-	Low
CZV	High	M-High		High	KEN	M-Low	Low	M-High	M-High
CZE	High	M-High	-	High	PPK	M-High	M-High	M-High	High
COG	-	Low	M-High	M-Low	KWT	M-High	High	-	M-Low
DNK	High	High	-	High	KGZ	-	M-Low	-	Low
DJI	-	M-Low	Low	-	LAO	-	Low	M-High	M-Low

TABLE 5. Groups by education, income, settler mortality, and population density in 1500. Low means that the country average is below the $25th$ percentile, M-Low between the $25th$ and $50th$ percentiles, M-High between the $50th$ and $75th$ percentiles, and High above the $75th$ percentile.

Country	Educ	Income	Settler Mortality	Pop Density	Country	Educ	Income	Settler Mortality	Pop Density
LVA	High	M-High	-	M-Low	SLE	Low	M-Low	High	M-High
LBN	-	M-High	-	M-High	SGP	M-High	High	Low	Low
LSO	M-Low	Low	-	Low	SVK	High	M-High	-	High
LBR	Low	Low	High	M-High	SVN	High	-	-	M-High
LBY	Low	M-High	-	Low	SOM	-	Low	-	M-Low
LTU	High	M-High	-	M-Low	ZAF	M-High	M-High	Low	Low
MKD	-	-	-	High	ESP	M-High	High	-	High
MDG	-	Low	High	M-Low	LKA	M-High	M-Low	M-Low	High
MWI	M-Low	Low	-	Low	SDN	Low	Low	M-High	M-Low
MYS	M-High	M-High	Low	M-Low	SWZ	M-Low	M-High	-	Low
MLI	Low	Low	High	Low	SWE	High	High	-	M-Low
MRT	Low	Low	High	Low	CHE	High	High	-	High
MUS	M-High	M-High	Low	Low	SYR	M-Low	M-High	-	M-High
MEX	M-Low	M-High	M-Low	M-High	TWN	M-High	M-High	-	
MDA	High	M-Low	-	M-Low	TJK	High	M-Low	-	Low
MNG	-	M-Low	-	Low	TZA	-	Low	High	M-Low
MAR	-	M-Low	M-High	M-High	THA	M-High	M-Low	M-High	M-High
MOZ	Low	Low	-	M-Low	TGO	Low	Low	High	M-High
KOR	-	M-High	-		TTO	M-High	High	M-High	M-Low
NAM	-	M-High	-	Low	TUN	Low	M-Low	Low	M-High
NPL	Low	Low	-	High	TUR	M-Low	M-High	-	High
NLD	High	High	-	High	TKM	-	M-High	-	Low
NZL	High	High	Low	Low	ARE	M-Low	High	-	M-Low
NIC	M-Low	M-Low	M-High	M-Low	UGA	Low	Low	High	M-High
NER	Low	Low	High	Low	GBR	High	High	Low	High
NGA	-	Low	High	M-High	UKR	-	M-High	-	M-Low
NOR	High	High	-	M-Low	URY	M-High	M-High	M-Low	
OMN	-	High	-	M-Low	USA	High	High	Low	Low
PAK	Low	Low	Low	High	UZB	-	M-High	-	Low
PAN	M-High	M-High	M-High	M-Low	VEN	M-Low	High	M-Low	Low
PNG	Low	M-Low	M-High	M-Low	VNM	M-Low	M-Low	M-High	M-High
PRY	M-High	M-Low	M-Low	Low	VNS	-	Low	M-High	M-High
PER	M-High	M-High	M-Low	M-Low	YEM	-	M-Low	-	M-High
PHL	M-High	M-Low	-	M-Low	YEA	-	Low	Low	M-High
POL	High	M-High	-	High	YEP	-	M-High	-	M-High
PRT	M-Low	M-High	-	High	YUG	M-High	M-High	-	High
ROM	High	Low	-	High	ZMB	M-Low	Low	-	Low
RUS	High	M-High	-	M-Low	ZWE	M-Low	M-Low	-	Low
RWA	Low	Low	High	High					
SAU	-	High	-	Low					
SEN	Low	M-Low	M-High	M-High					

TABLE 6. Groups by education, income, settler mortality, and population density in 1500. Low means that the country average is below the $25th$ percentile, M-Low between the $25th$ and $50th$ percentiles, M-High between the $50th$ and $75th$ percentiles, and High above the $75th$ percentile. (Cont.)

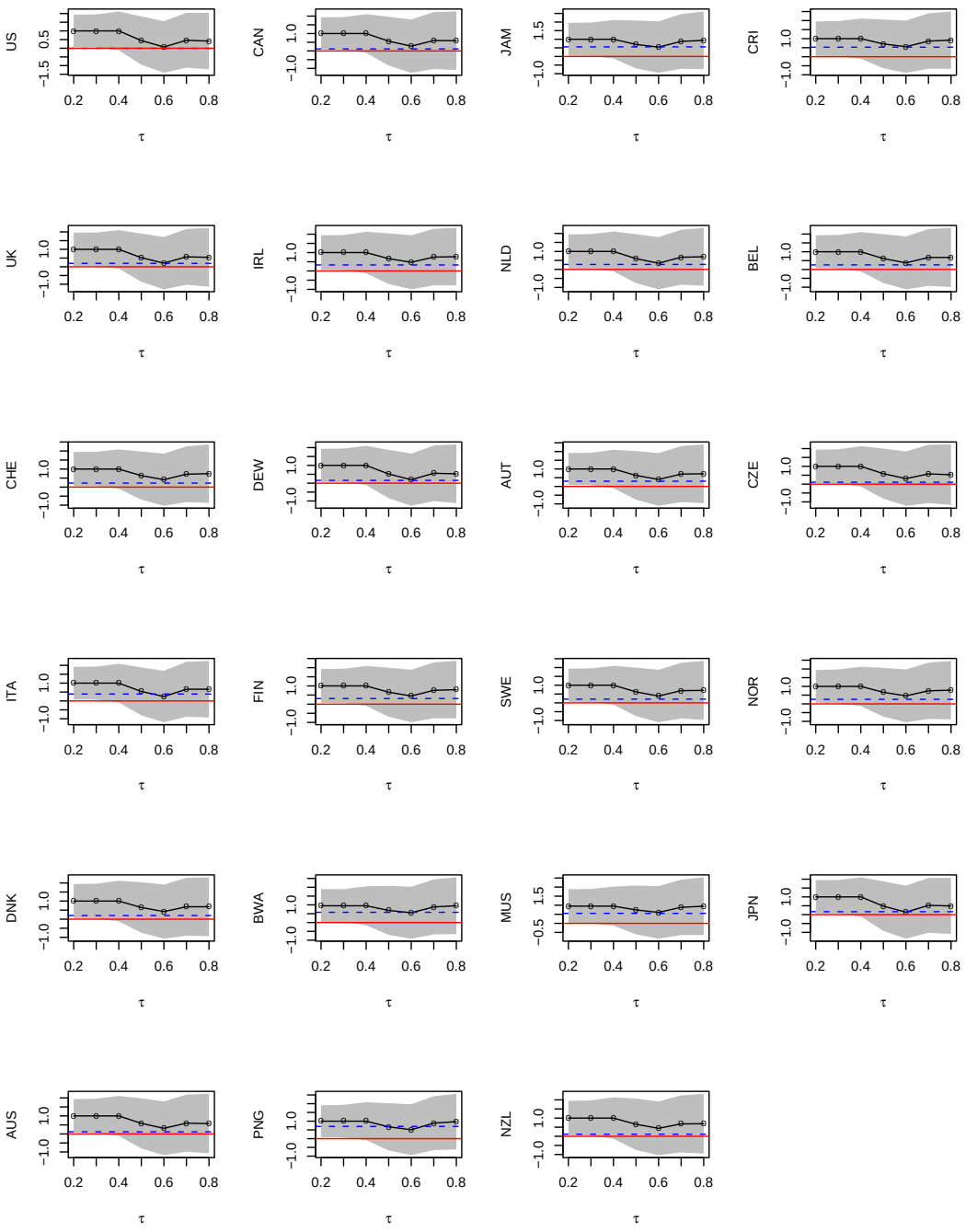

FIGURE 1. Country effect estimates in a model for the Polity IV measure of democracy. The figure presents results for countries with at least one individual effect significant at 1 percent, in a model that includes log of GDP at t-1, log of population, and schooling. The continuous line with dots represents the quantile regression country effects estimates and the dashed blue line the mean fixed effect. The shaded grey area represents a pointwise confidence interval.

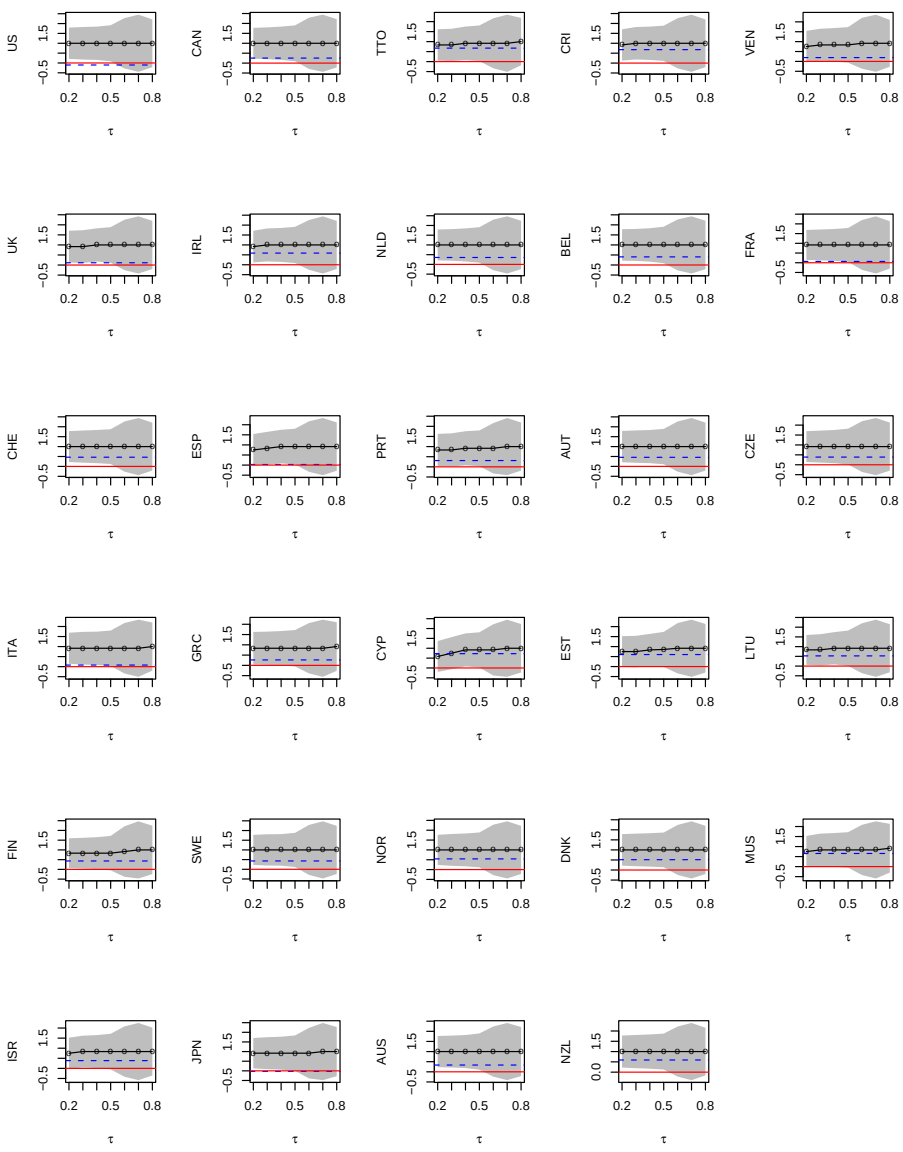

FIGURE 2. Country fixed effect estimates in a model for the Freedom House measure of democracy. The figure presents results for countries with at least one individual effect significant at 1 percent, in a model that includes log of GDP at t-1 and log of population. The continuous line with dots represents the quantile regression fixed effects estimates and the dashed blue line the mean fixed effect. The shaded grey area represents a pointwise confidence interval.

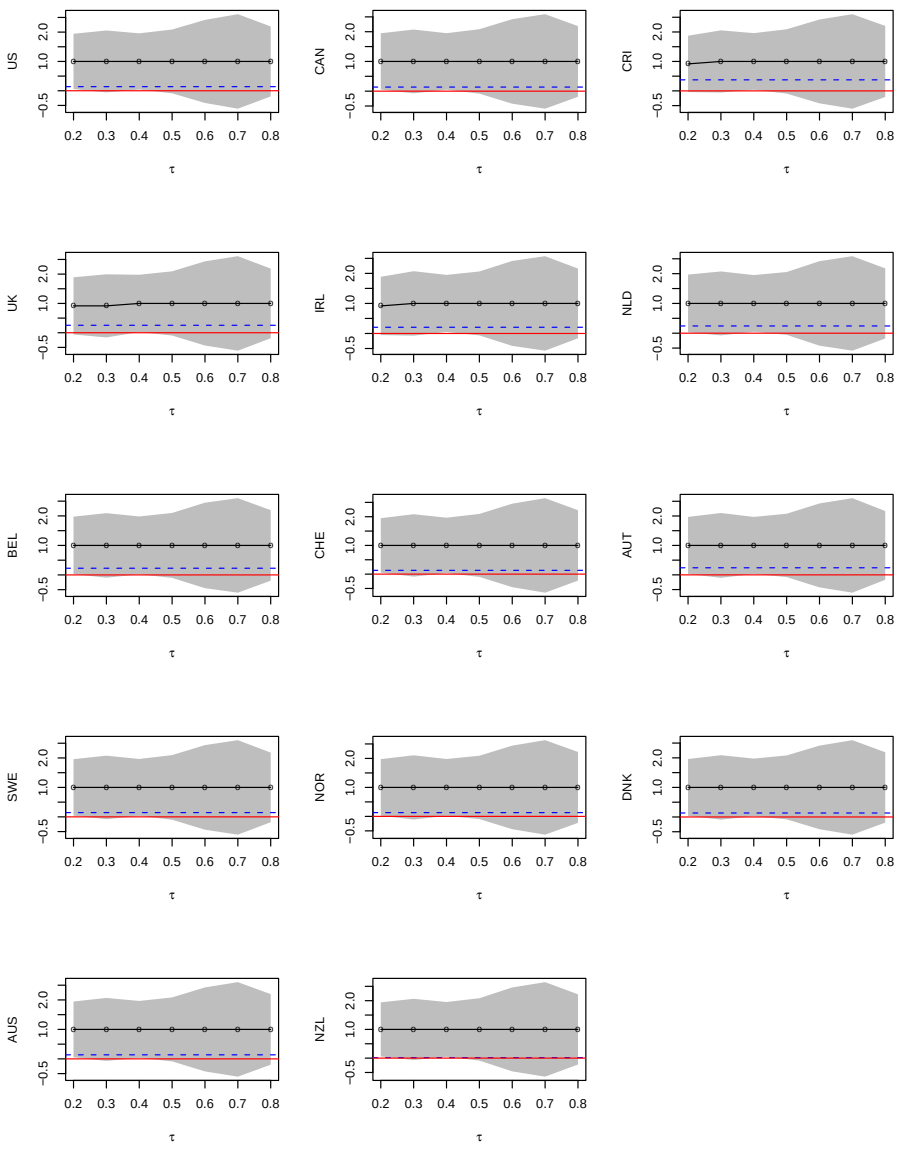

FIGURE 3. Country fixed effect estimates in a model for the Freedom House measure of democracy. The figure presents results for countries with at least one individual effect significant at 1 percent, in a model that includes log of GDP at t-1, log of population, and schooling. The continuous line with dots represents the quantile regression fixed effects estimates and the dashed blue line the mean fixed effect. The shaded grey area represents a pointwise confidence interval.

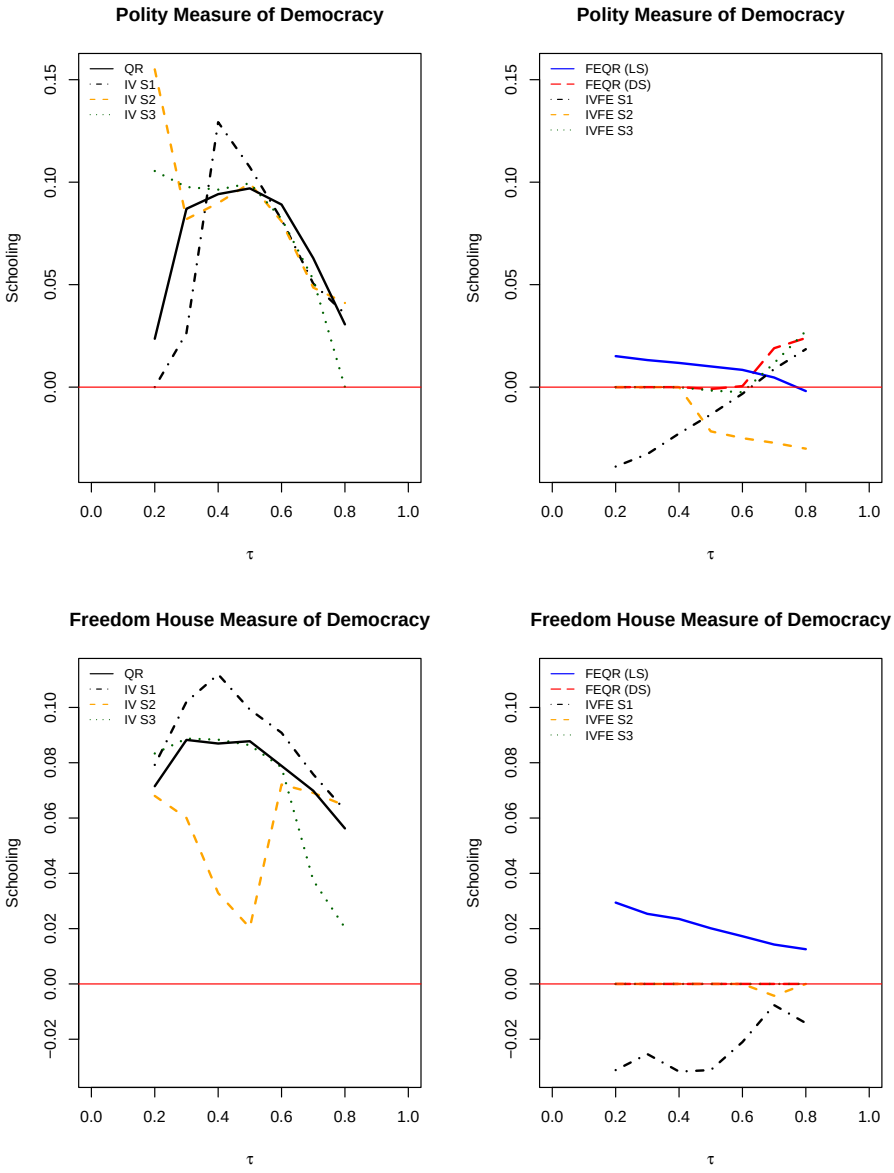

FIGURE 4. Quantile regression estimates of the effect of education on democracy considering the full sample of countries. The panels show point estimates obtained by using quantile regression for the pooled data (QR), instrumental variable with and without fixed effects (IVFE, IV), and quantile regression with fixed effects assuming that the individual effects are either location shifts (LS) or distributional shifts (DS). The instrument sets (S1-S3) are described in Table 3.

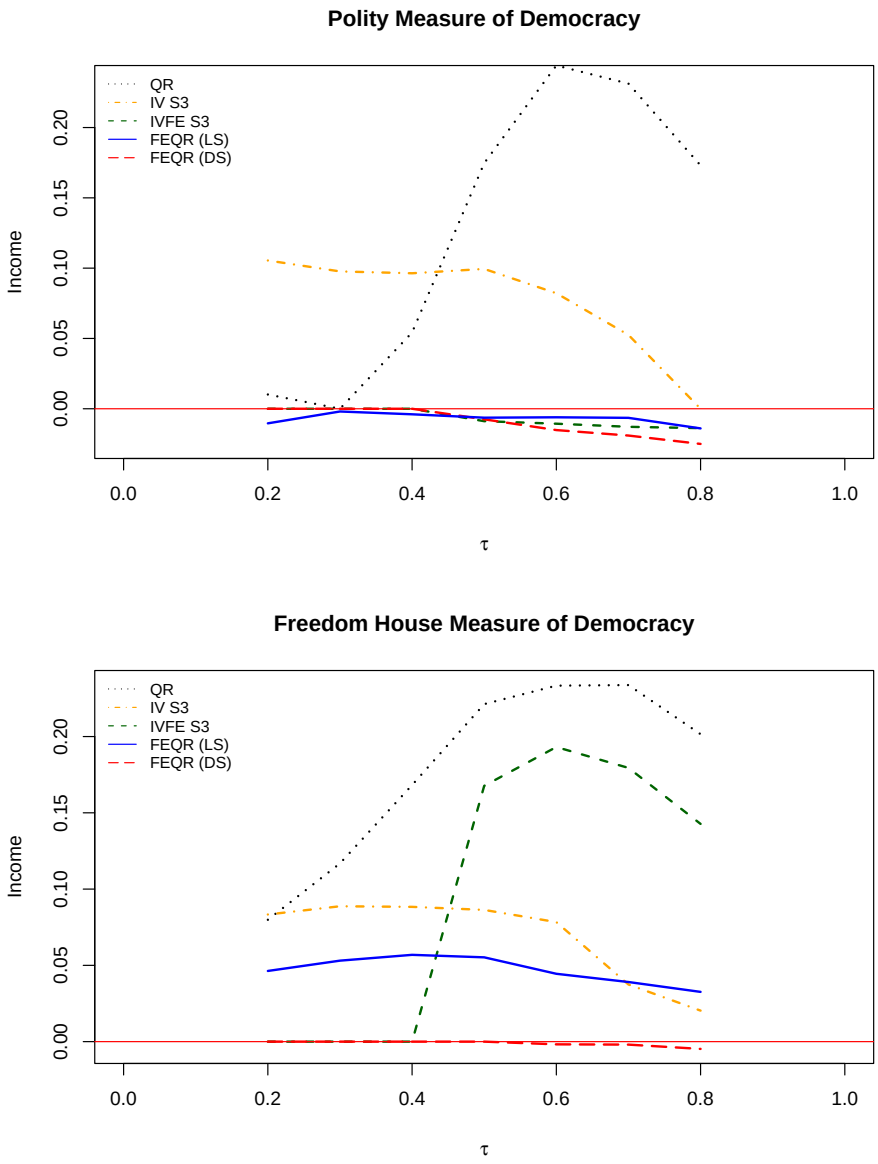

FIGURE 5. More evidence on the distributional effect of income on democracy. The sample includes countries with positive standard deviation of democracy. The panels show point estimates obtained by using quantile regression for the pooled data (QR), instrumental variable with and without fixed effects (IVFE, IV), and quantile regression with fixed effects assuming that the individual effects are either location shifts (LS) or distributional shifts (DS). The instrument set S2 includes trade-weighted world income as in Acemoglu et. al. (2008).

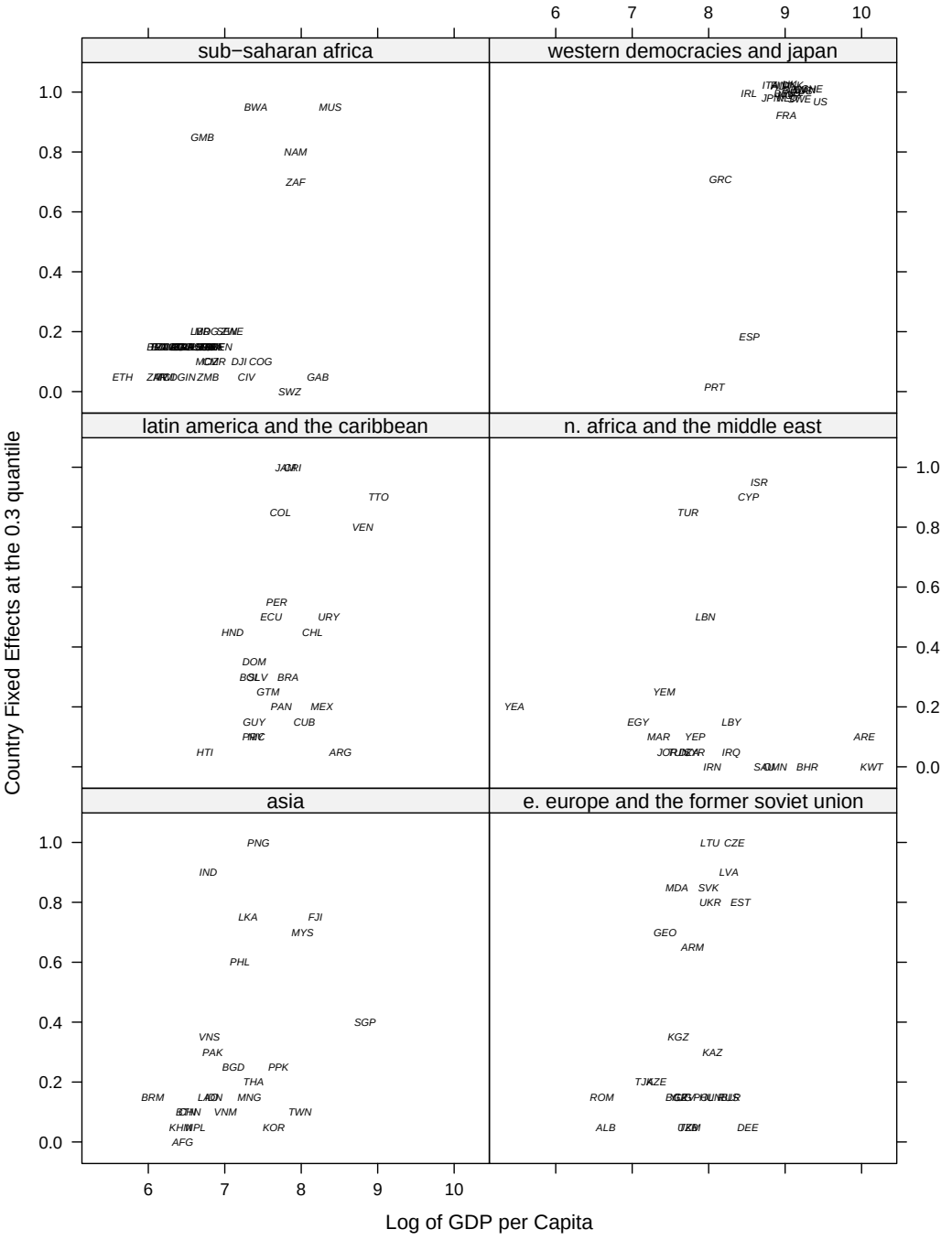

FIGURE 6. Scatterplot of country effects and income. The country effects were estimated at the 0.3 quantile of the conditional distribution of democracy, in a linear model that includes log of DGP at $t-1$ and log of population. The dependent variable democracy is constructed based on the Polity measure. See the Appendix for data definitions.

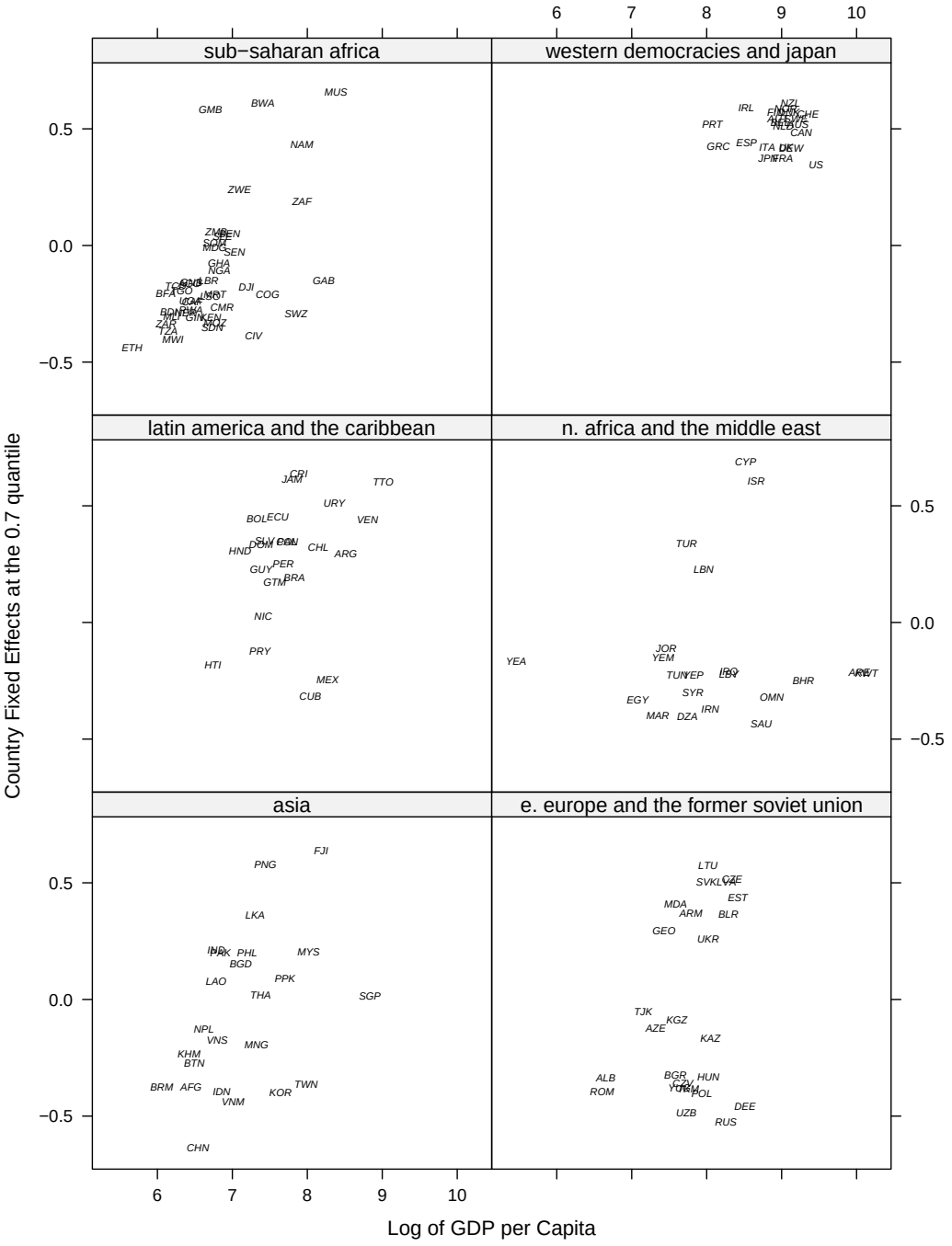

FIGURE 7. Scatterplot of country effects and income. The country effects were estimated at the 0.7 quantile of the conditional distribution of democracy, in a linear model that includes log of DGP at $t-1$ and log of population. The dependent variable democracy is constructed based on the Polity measure. See the Appendix for data definitions.

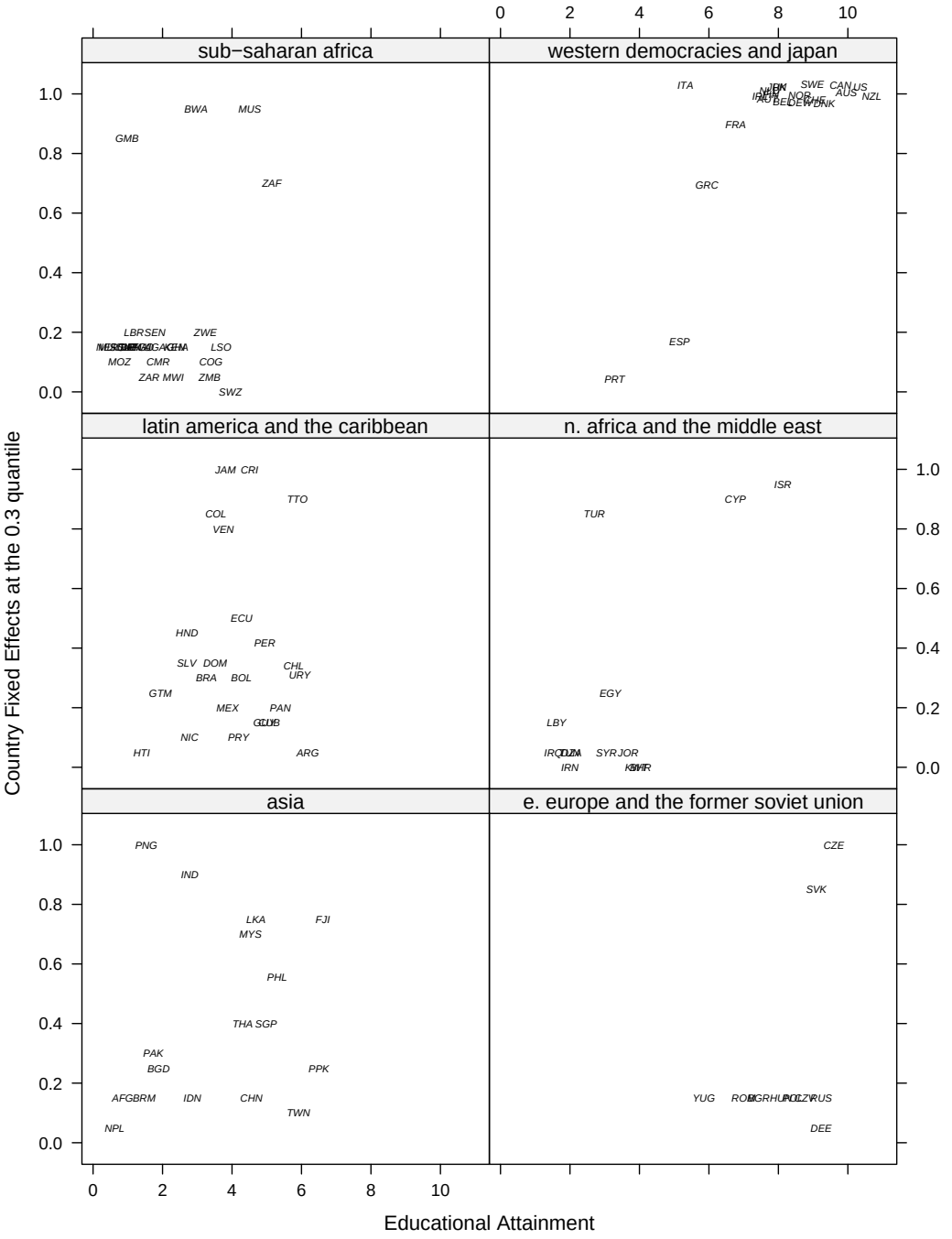

FIGURE 8. Scatterplot of country effects and average years of schooling. The country effects were estimated at the 0.3 quantile of the conditional distribution of democracy, in a linear model that includes log of DGP at $t-1$, log of population, and schooling. The dependent variable democracy is constructed based on the Polity measure. See the Appendix for data definitions.

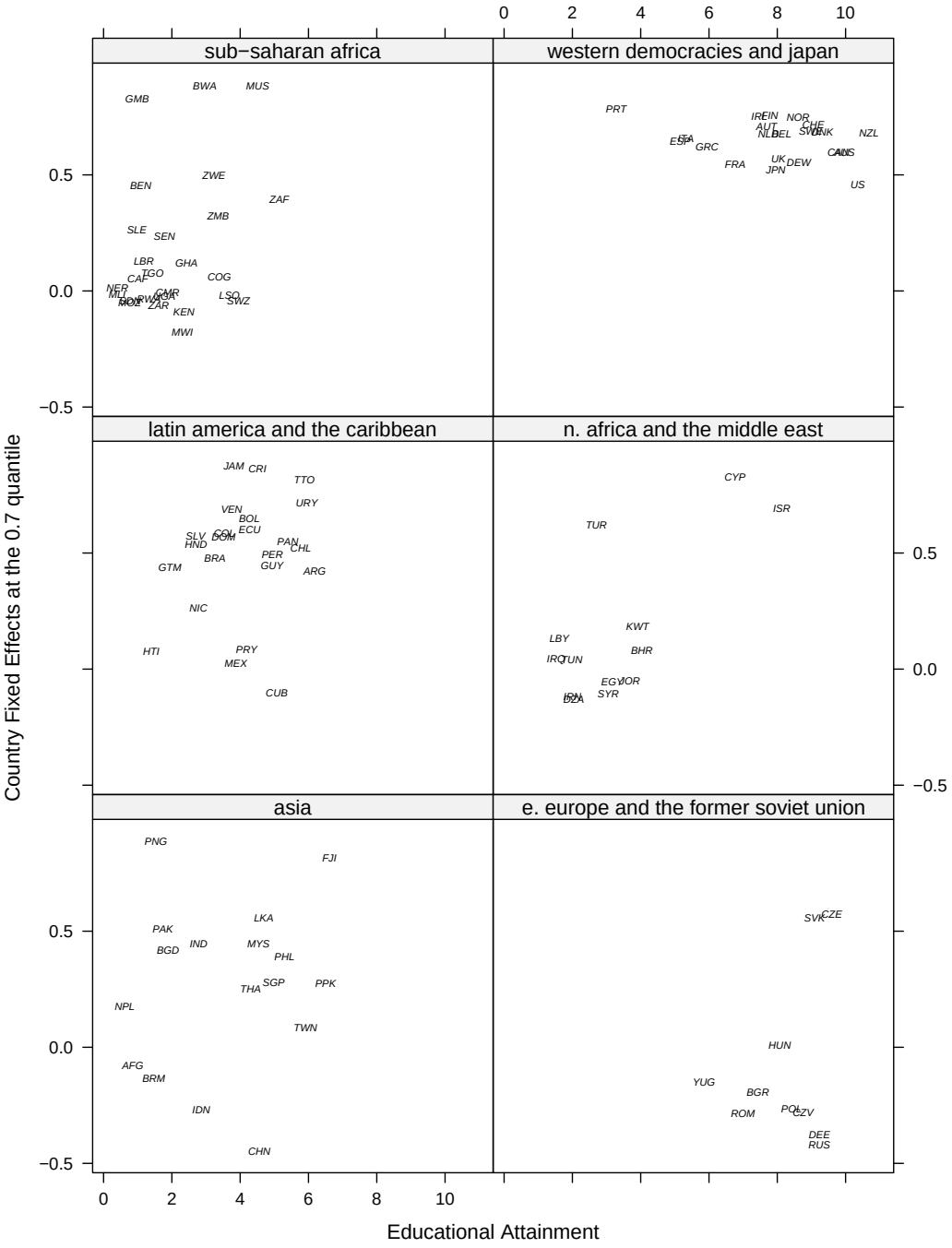

FIGURE 9. Scatterplot of country effects and average years of schooling. The country effects were estimated at the 0.7 quantile of the conditional distribution of democracy, in a linear model that includes log of DGP at $t-1$, log of population, and schooling. The dependent variable democracy is constructed based on the Polity measure. See the Appendix for data definitions.