Discussion Papers
Department of Economics
University of Copenhagen

No. 17-16

Instrumental Variables in the Long Run

Gregory Casey and Marc Klemp

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Instrumental Variables in the Long Run*

Gregory Casey Marc Klemp

August 8, 2017

Abstract

In the study of long-run economic growth, it is common to use historical or geographical variables as instruments for contemporary endogenous regressors. We study the interpretation of these conventional instrumental variable (IV) regressions in a general, yet simple, framework. Our aim is to estimate the long-run causal effect of changes in the endogenous explanatory variable. We find that conventional IV regressions generally cannot recover this parameter of interest. To estimate this parameter, therefore, we develop an augmented IV estimator that combines the conventional regression with a separate regression estimating the degree of persistence in the endogenous regressor. Importantly, our estimator can overcome a particular violation of the exclusion restriction that can arise when there is a time gap between the instrument and the endogenous explanatory variable. We apply our results to estimate the long-run effect of institutions on economic performance and the long-run effect of Protestantism on human capital accumulation. In both cases, we find economically significant long-run effects that are smaller than those in the existing literature, demonstrating that our results have important quantitative implications for the field of long-run economic growth. We also use our framework to examine related empirical techniques. We find that two prominent regression methodologies – using gravity-based instruments for trade and including ancestry-adjusted variables in linear regression models – have related issues of interpretation. In the latter case, this problem can be overcome by including both unadjusted and adjusted measures in the regression model.

Keywords Long-Run Economic Growth, Instrumental Variable Regression

JEL Classification Codes C10, C30, O10, O40

---

*We wish to thank Daron Acemoglu, Mario Carillo, Kenneth Chay, Carl-Johan Dalgaard, Melissa Dell, Andrew Dickens, Mette Ejrnæs, Diego Focanti, Andrew Foster, Raphael Franck, Oded Galor, Philipp Ketz, Daniel le Maire, Stelios Michalopoulos, Ömer Özak, Jim Robinson, Gerard Roland, Sanjay Singh, Tim Squires, Uwe Sunde, David Weil, Ben Zou and participants at the Brown University Macro Lunch, University of Copenhagen Economic Growth Mini Workshop, NEUDC 2015, and the 2016 Zeuthen Conference for valuable comments. The research of Klemp is funded by the Carlsberg Foundation and by the Danish Research Council reference no. 1329-00093 and reference no. 1327-00245. Gregory Casey (Gregory_Casey@brown.edu): Department of Economics, Brown University, 64 Waterman St., Providence, RI 02912, USA. Marc Klemp (marc_klemp@brown.edu): Department of Economics and Population Studies and Training Center, Brown University, 64 Waterman St., Providence, RI 02912, USA, and Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26, DK-1353 Copenhagen K, Denmark. Any errors and all opinions are our own.

1 Introduction

A growing literature examines the determinants of long-run economic development (Spolaore and Wacziarg, 2013; Nunn, 2014; Ashraf and Galor, 2016). In this literature, it is common to use historical or geographic instruments for contemporary endogenous regressors in instrumental variables (IV) regressions (e.g., Acemoglu et al., 2001; Easterly, 2007; Becker and Woessmann, 2009; Tabellini, 2010).[1] In this paper, we study the interpretation of these conventional IV regressions, develop an augmented estimator for long-run causal effects, and apply our findings to two prominent studies in the field of long-run economic growth. We also discuss the implications of our findings for related empirical techniques, including spatial regression discontinuity, estimation using gravity-based instruments for trade, and ordinary least squares (OLS) regressions with ancestry-adjusted variables.

Despite the prominence of instrumental variable regressions with historical instruments and contemporary endogenous regressors, specific interpretations are rarely attached to the estimated coefficients. We provide a general, yet simple, framework for interpreting these regressions that is consistent with the existing literature. Our parameter of interest is the 'long-run effect' of historical changes in the endogenous explanatory variable on the contemporary dependent variable. This is the parameter that would be estimated by standard IV analysis if the endogenous regressor was measured at the time of the initial impact of the instrument. This is also the parameter that provides information about the long-run consequences of policy interventions or historical events. We find that IV regressions in which the endogenous regressor is measured later in time estimate the ratio of the long-run effect and the persistence of the endogenous variable. Here, we use 'persistence' to denote the causal effect of the historical (i.e., related to the time period of the initial impact of instrument) level of the endogenous variable on the contemporary (i.e., related to the time period of the outcome variable) level of the endogenous variable. Our analysis, therefore, shows that accounting for the persistence in the endogenous regressor is important for estimating long-run causal effects when the endogenous regressor is observed after the initial impact of the instrument.

Based on these results, we develop an augmented estimator for the long-run causal effects under common data availability constraints. Specifically, we consider the case where the endogenous explanatory variable is not measured at the time of the original impact of the instrument. This new estimator can be implemented by jointly estimating two equations using a single instrument.[2] One equation estimates the conventional regression. The other estimates the persistence of the endogenous variable between two intermediate points in time. The second equation can then be used to correct the bias from the conventional regression.

We show that our approach can estimate the long-run effect in the presence of certain violations of the exclusion restriction that we argue are common in the field of long-run growth. Furthermore,

[1]This technique is still popular in the literature (e.g., Becker et al., 2010; Naritomi et al., 2012; Auer, 2013; Ang, 2013; Acemoglu et al., 2014; Gorodnichenko and Roland, 2011, 2017).

[2]The system can be estimated using, for example, multiple-equation Generalized Method of Moments (GMM) or stacked 2SLS regressions.

we show that, in the presence of these violations, it is not possible to estimate the contemporaneous causal relationship between the endogenous regressor and the outcome variable. Thus, a key aspect of our study is to demonstrate how to extract an interesting economic parameter under violations of the exclusion restriction.

We use our results to estimate the long-run effect of institutions on economic performance and the long-run effect of culture on human capital formation. We base our estimations on the instrumental variable strategies and data of two prominent articles within the field of long-run economic growth. We start by applying our method to cross-national data on institutions, building on the work of Acemoglu et al. (2001). In our preferred specification, a change in constraints on executive power in 1800 from the lowest to the highest possible score on the Polity IV index leads to a 1.3 standard deviation change in 1990 income per capita. While sizable, this effect is 45 percent smaller than the coefficient generated by the conventional IV regression, indicating that our augmented estimator is quantitatively important.[3]

In our second application, we examine the effect of the spread of Protestantism on human capital formation in Prussia, building on the work of Becker and Woessmann (2009).[4] We find that an increase in the Protestant share of a Prussian county in 1517 from 0 to 100 percent raises literacy in 1870 by 1.5 standard deviations.[5] This effect is eight percent smaller than the conventional IV estimate. Thus, once again, we find that accounting for persistence with our augmented regression technique is quantitatively important.

After presenting our core results, we provide a broader discussion that reviews practical implications for empirical researchers and examines the implications of our framework for other empirical techniques used in the study of long-run economic growth. Gravity-based instruments for trade or migration are also vulnerable to the violations of the exclusion restrictions discussed in this paper, but our augmented approach cannot correct this problem to estimate the long-run effect. At the same time, OLS regressions with ancestry-adjusted variables have related issues of interpretation, which can be overcome by including both adjusted and unadjusted measures in the regressions. Spatial regression discontinuity and collection of historical data can help overcome the issues of identification raised in this paper and estimate long-run effects without employing our augmented technique.

Our results have important implications for the field of long-run economic growth. First, we provide an interpretation for IV regressions with historical instruments and contemporary endogenous regressors. Second, we provide a new procedure that enables researchers to estimate the long-run effect of potential determinants of economic performance and related outcomes. Third, using our new analytic results and empirical technique, we generate updated estimates of the impact of institutions and culture on economic development.[6] Finally, we use the insights from our

[3]We also use panel data on institutions to validate key assumptions in our new method.

[4]This work, in turn, builds the influential hypothesis of Weber (1958) about the relationship between Protestantism and economic development, but stresses the effect of religion on human capital, rather than work ethic.

[5]Following the original paper, we also examine the effect of Protestantism on distance to the nearest school. We find very similar results in this case.

[6]For related work on the long-run impact of institutions using alternate empirical techniques, see Banerjee and Iyer (2005), Dell (2010), Bruhn and Gallego (2012), and Michalopoulos and Papaioannou (2013). For related work

analysis to provide a new perspective on a wide range of empirical techniques used in the long-run growth literature.

More generally, our analysis demonstrates the importance of considering the underlying data generating process when interpreting long-run growth regressions. In particular, we show how considering even a simple representation of the underlying dynamic relationships can affect the interpretation of commonly used regression techniques. In this way, our results are closely related to work by Acemoglu (2010) and Deaton (2010a,b), who also stress the importance of utilizing theory to make sense of empirical results in economic development.

In Section 2, we present our framework and main analytic results. Section 3 presents our empirical application. Section 4 examines practical implications for applied researchers and extends the discussion to other empirical techniques. Section 5 concludes.

2 Framework and Analytic Results

Problems of omitted variables and reverse causality are abundant in the expanding literature on the fundamental determinants of economic growth. To estimate causal effects, therefore, researchers often use historical or geographical variables as instruments for contemporary determinants of economic development. As our analysis shows, however, the time lag between the instrument and the endogenous variable complicates the interpretation of the regression coefficient.

In Section 2.1, we provide a general framework for interpreting instrumental variable regressions when the instrument precedes the endogenous regressor in time. Our parameter of interest is the 'long-run effect' of historical changes in the endogenous explanatory variable on the contemporary dependent variable. This parameter tells us about the long-run implications of a given policy or historical event, which are of fundamental importance in this literature. We use our framework to derive the relationship between our parameter of interest and the coefficient from a conventional IV regression. The difference depends on the 'persistence' in the endogenous variable, which we define as the causal effect of changes in historical level of the endogenous variable on the contemporary level. Throughout our analysis, we use 'historical' to refer to the time period in which the instrument first exerts an impact on the dependent variable and 'contemporary' to indicate the time in which the dependent variable is measured.

Our framework explicitly accounts for certain violations of the exclusion restriction that are important in the field of long-run growth. The existence of these violations does not affect our core results. They do, however, prevent the conventional IV regression from estimating the contemporaneous causal relationship between the endogenous and outcome variable. Thus, a key aspect of our study is to demonstrate how to estimate interesting economic parameters under violations of the exclusion restriction.

Section 2.2 builds on the results from Section 2.1 by demonstrating how to augment conventional IV regressions in order to recover our parameter of interest. Our augmented estimator extracts

on the long-run impact of culture using alternate empirical techniques, see Guiso et al. (2006), Alesina et al. (2011), and Alesina and Giuliano (2013).

the long-run causal effect of the endogenous variable by explicitly estimating the persistence of the endogenous variable using observations at two intermediate points in time. Our method can be implemented with multiple-equation GMM or stacked 2SLS regression using a single instrument.[7]

2.1 Interpreting IV regressions in the long-run growth literature

Figure 1 provides a representation of our framework. We start by just considering the top row (i.e. we ignore A). Our endogeneous explanatory variable of interest is X, and Y is the dependent variable. The explanatory variable, X, is time-varying. We use the subscript H to denote the historical time period and C to denote the contemporary period. Throughout our analysis, we use 'historical' to refer to the time period in which the instrument first exerts an impact on X and 'contemporary' to indicate the time in which Y is measured. We assume that Z would be a valid instrument for X_H, but that X_H is unobserved. This is a common data availability constraint in the long-run growth literature. We are interested in examining the causal effect of X_H on Y, which we refer to as the long-run effect of X on Y. A data generating process of this form is usually implicitly assumed to underlie regressions of this type. Without the presence of A, Z is a valid instrument for X_C, and it is possible to estimate the causal effect of X_C on Y with a standard instrumental variables regression. All of our key results will also hold in this basic setting.

We believe, however, that the top row of Figure 1 provides an incomplete picture of the underlying dynamics in most cases. Our reasoning is as follows: if there are good reasons to expect that X_C affects Y in the contemporary period, then X_H should in general also affect Y in the historical period. In that case, if there is persistence in Y — or if the factors through which X_H affects historical values of Y are persistent — then there will be a causal effect of X_H on Y that is not intermediated by X_C. We represent this link using the variable A. In most applications, it is unlikely that all components of A are observed.[8] Thus, we assume that A is unobserved. We will refer to A as an 'alternative channel.'[9]

At first, the existence of an alternative channel may appear problematic because it violates the usual exclusion restriction. As we will show, however, the presence of A will not inhibit our ability to estimate the parameter of interest. Indeed, a key insight generated by our framework is that, in the presence of an A variable, our parameter of interest (i.e., the long-run effect) remains the only avenue to learn about the causal effect of X on Y. Since it is more realistic, therefore, we focus on the case in which there exists an A variable — though we stress again that its inclusion does not affect our main result regarding the method to estimate the long-run effect.[10] In particular, the validity of our estimator of the long-run effect does not depend on the existence of an A variable. We model these alternative channels in a reduced form manner, but a wide range of alternate specifications can be re-written in this form.

[7] A Stata program for application of our method is available upon request and will be released with the paper.

[8] Compared to our framework, observing part, but not all, of the A variable would further complicate the interpretation of the regression. Moreover, the components of A that are observed are unlikely to be uncorrelated with the error term in most applications.

[9] For the remainder of the paper, when we refer to an A variable or an alternative channel, we focus on the case where $\gamma \neq 0$ and $\beta_2 \neq 0$.

[10] Appendix Section A.1 analyzes the case without A.

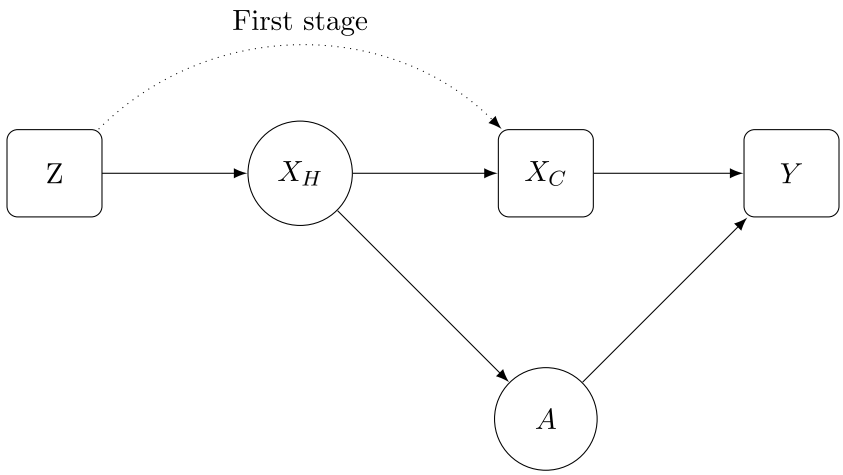

Figure 1: Causal diagram of equations (1)–(4) and the first stage in a conventional 2SLS regression. Rectangular nodes represent observed variables and circular nodes represent unobserved variables. The dotted line represents the first stage in a conventional 2SLS estimation.

To fix ideas, it is helpful to consider a particular example. Our system is a generalization of the data generating process presented in Acemoglu et al. (2001). In their framework, Z is settler mortality, Y is income per capita, and X is institutional quality. Compared to their formal presentation of the underlying model, we include the existence of the A variable, which is consistent with the empirical findings and interpretation presented in their paper.[11] The A variable could be physical or human capital, technology, or culture.

Equations (1)–(4) represent the data generating process algebraically:

$$\begin{aligned} X_{H,i} &= \kappa_{X_H} + \psi Z_i + \varepsilon_{H,i}, & (1)\\ X_{C,i} &= \kappa_{X_C} + \delta X_{H,i} + \varepsilon_{X_C,i}, & (2)\\ A_i &= \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i}, & (3)\\ Y_i &= \kappa_Y + \beta_1 X_{C,i} + \beta_2 A_i + \varepsilon_{Y,i}. & (4) \end{aligned}$$

In standard settings, instrumental variables are used to estimate the contemporaneous causal effect of X, $\frac{\partial Y}{\partial X_C} = \beta_1$. Our parameter of interest is the long-run causal effect of X, $\eta \equiv \frac{\partial Y}{\partial X_H} = \beta_1\delta + \beta_2\gamma$. Another parameter that plays a key role in our framework is $\frac{\partial X_C}{\partial X_H} = \delta$, which measures the 'persistence' of historical changes in X. If $\delta > 1$, then the endogenous variable diverges from its original path following a shock. If $\delta < 1$, then it converges back to its original path, and shocks eventually die out.

[11] In particular, Acemoglu et al. (2001) find that historical institutions exert an impact on contemporary income independently of contemporary institutions. Their interpretation of these results is in line with our equations: "In some specifications, the overidentification tests using measures of early institutions reject at that 10-percent level (but not at the 5-percent level). There are in fact good reasons to expect institutions circa 1900 to have a direct effect on income today (and hence the overidentifying tests to reject our restrictions): these institutions should affect physical and human capital investments at the beginning of the century, and have some effect on current income levels through this channel" (fn 31, p. 1393).

The conventional IV regression of Y on X_C with Z as an instrument yields

$$\text{plim}\,\hat{\beta}_1^{IV} = \beta_1 + \frac{\beta_2\gamma}{\delta} = \frac{\eta}{\delta}. \tag{5}$$

Thus, the conventional 2SLS coefficient is consistent for the ratio of the long-run effect and the persistence of the endogenous variable. This has an intuitive interpretation in that a one-unit change in X_C is associated with a δ^{-1} unit change in X_H. In other words, the conventional regression measures the long-term impact of an increase in X_H that corresponds to a one-unit rise in X_C. A large conventional regression coefficient, therefore, may, indicate either a large impact of X_H or low persistence in X.

The algebra also indicates that, in the presence of an A variable, it is not possible to recover the contemporaneous causal relationship, β_1. The IV coefficient overestimates η when X converges to its original path after a shock (i.e., $\delta < 1$) and underestimates the effect when X diverges over time following a shock (i.e., $\delta > 1$). The two are equal only in the knife-edge case where $\delta = 1$. We refer to this condition as X being 'perfectly persistent.' In light of these results, it is apparent that in the absence of information on the persistence of the endogenous variable, the conventional IV coefficient is uninformative about the long-run effect of X on Y. As demonstrated in Section A.1 in the appendix, the relationship between the regression coefficient and η is unchanged if the A variable is excluded from the system.

These results suggest that we could recover η by multiplying the conventional IV coefficient by δ or by using X_H, rather than X_C, in the regression. In most applications in long-run economic growth, however, X_H is not observed. Thus, we need to combine the cross-sectional regression with an estimate of δ. In the next subsection, we demonstrate how to estimate η in this manner.

2.2 Estimating the Long-Run Effect

In this section, we demonstrate how to estimate η when X_H is not observed. In order to estimate δ, we make use of measures of X at two intermediate points in time. Thus, our framework extends that of the previous section by allowing for more than two time periods:

$$X_{t,i} = \kappa_{X_t} + \delta X_{t-1} + \varepsilon_{X_t,i},\ \forall\ t = 1 \ldots C,\ t \neq H, \tag{6}$$

$$X_{H,i} = \kappa_{X_H} + \delta X_{H-1,i} + \psi Z_i + \varepsilon_{X_H,i}, \tag{7}$$

$$A_i = \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i}, \tag{8}$$

$$Y_i = \kappa_Y + \beta_1 X_{C,i} + \beta_2 A + \varepsilon_{Y,i},. \tag{9}$$

The timing of this set-up is such that X initially follows a simple law of motion given by (6). Then, in some period H, X is shocked by Z. After the shock, X continues to follow the original law of motion. These assumptions allow us to infer the relationship between X_C and X_H even when the latter is not observable.[12]

[12] This result can be generalized to other functional forms or a time-varying δ, provided that sufficiently strong assumptions are made.

We start by solving for the relationship between values of X_T and X_{T-Q}, which we will use to estimate the degree of persistence, $\frac{\partial X_C}{\partial X_H}$. To do so, we simply apply (6) recursively:

$$X_{T,i} = \kappa_T + \delta X_{T-1,i} + \varepsilon_{X_T,i} = \sum_{k=0}^{Q-1} \delta^k \kappa_{X_{T-k}} + \delta^Q X_{T-Q,i} + \sum_{k=0}^{Q-1} \delta^k \varepsilon_{X_{T-k},i}. \quad (10)$$

Now, consider the IV regression equation:

$$X_{T,i} = a_0 + a_1 X_{T-Q,i} + a_{2,i}, \quad (11)$$

where Z is an instrument for $X_{T-Q,i}$. There is no violation of the exclusion restriction in this case, and according to (10), the estimation yields:

$$\text{plim}\, \hat{a}_1 = \delta^Q. \quad (12)$$

Next, turn to the relationship between X and Y. A little algebra yields

$$Y_i = \tilde{\beta}_0 + (\beta_1 \delta^{C-H} + \beta_2 \gamma) X_{H,i} + \tilde{\epsilon}_i, \quad (13)$$

where $\tilde{\beta}_0 = \kappa_{X_C} + \beta_1 \sum_{k=0}^{C-H-1} \delta^k \kappa_{X_{T-k}} + \beta_2 \kappa_A$ and $\tilde{\epsilon}_i = \beta_1 \sum_{k=0}^{Q-1} \delta^k \varepsilon_{X,T-k} + \varepsilon_{X_C,i} + \beta_2 \varepsilon_{A,i}$. It follows immediately that $\eta \equiv \frac{\partial Y}{\partial X_H} = \beta_1 \delta^{C-H} + \beta_2 \gamma$. Now, consider the conventional IV regression model,

$$Y_i = b_0 + b_1 X_{C,i} + b_{2,i}, \quad (14)$$

where Z_i is an instrument for X_C. Similar to our results from Section 2, this regression yields

$$\text{plim}\, \hat{b}_1 = \frac{\beta_1 \delta^{C-H} + \beta_2 \gamma}{\delta^{C-H}} = \frac{\eta}{\delta^{C-H}}. \quad (15)$$

To solve for η, we simply combine the results from estimating equations (11) and (14):

$$\text{plim}\, \hat{a}_1 = \delta^Q \;\Rightarrow \delta = (\text{plim}\, \hat{a}_1)^{\frac{1}{Q}} \quad (16)$$

and

$$\text{plim}\, \hat{b}_1 = \frac{\eta}{\delta^{C-H}} \;\Rightarrow \eta = (\text{plim}\, \hat{b}_1)\delta^{C-H} = (\text{plim}\, \hat{b}_1)(\text{plim}\, \hat{a}_1)^{\frac{C-H}{Q}}. \quad (17)$$

To estimate η, we first estimate equations (11) and (14) via instrumental variables in order to obtain $\hat{b}_1$ and $\hat{a}_1$.[13] Then, we combine the two regression coefficients using the nonlinear function in (17). To construct confidence intervals, we apply the delta method.

It is worth noting two key assumptions in our framework. First, we assume that the effect of X_H on X_C is linear. Second, we assume that δ is constant over time. The first of these assumptions

[13] These equations can be jointly estimated, e.g., via stacked 2SLS regressions or multiple-equation instrumental variable GMM. It should be noted that it is possible to include control variables in these models.

can be examined whenever our method can be applied, i.e., whenever measures of the endogenous variable is available at two points in time. The second assumption can be examined whenever measures are available at least for three points in time. In the applications below, we investigate the validity of these assumptions, using panel data when possible.

Overall, in Section 2.1, we provided a general, yet simple, framework for interpreting IV regression coefficients with historical instruments and contemporary endogenous regressors. We found that the regression coefficient estimates the long-run impact of changing historical conditions divided by the persistence of the endogenous variable. Furthermore, we found that this is true even under certain violations of the exclusion restriction, namely in the presence of an omitted A variable. In Section 2.2, we provided an augmented estimator that uses multiple regressions to estimate the long-run impact of changing historical conditions.

3 Applications

In this section, we apply our findings to two important questions in long-run economic growth. In particular, we estimate the impact of institutions on income per capita and the impact of religion on human capital accumulation.

3.1 Institutions and Income per Capita

In our first application, we examine the effect of institutions on economic performance, following Acemoglu et al. (2001). We choose this application for several reasons. First, the analysis by Acemoglu et al. (2001) is likely the most prominent paper using historical instruments for contemporary endogenous regressors and many important papers in the institutions literature have followed suit.[14] Moreover, unlike many subsequent papers using this empirical technique, Acemoglu et al. (2001) provide an explicit set of equations for interpreting their results, as well as a discussion of the role of past institutions. Our framework is consistent with their equations and discussion, making our new results immediately applicable in this context (see footnote 11). Finally, given the prominence of the institutions literature, much effort has gone into collecting measures of institutional characteristics of countries at different points in time. These data are essential in our empirical application and also allow for validation exercises.

3.1.1 Results

Our measure of institutional quality, 'Constraints on the Executive,' comes from the Polity IV dataset. It measures the limits to executive power on a seven point scale that increases in the level of constraints. This is a preferred measure of institutional quality used in the IV literature (Glaeser et al., 2004; Acemoglu et al., 2005). The outcome variable is the natural log of income per capita

[14] See for example Easterly and Levine (2003), Rodrik et al. (2004), and Acemoglu and Johnson (2005).

in 1990, and the instrument is settler mortality.[15] Since settler mortality may be correlated with region-specific factors, such as the disease environment or geography, that also affect contemporary income, we include controls for the log of the absolute value of latitude and World Bank region fixed effects.[16] Table A.7 in the appendix provides summary statistics.

We apply our new estimator from Section 2.2 to measure the long-run effect of institutions on economic development. To do so, we simultaneously estimate two equations via stacked 2SLS. First, we estimate the cross-sectional relationship between contemporary institutions and contemporary income per capita via equation (11). Second, we estimate the persistence of institutions via equation (14). Then, we use equation (17) to combine the results and extract the long-run effect of improving institutions. Both equations are estimated using settler mortality as an instrument, following Acemoglu et al. (2001).[17] As discussed in Section 2.1, the summary variable A captures the joint impact of several factors, including human and physical capital accumulation, technology, or persistence in income.

The timing of the initial shock is difficult to determine exactly and likely differs across countries. We take a conservative approach and use the year 1800. Using an earlier time period would only increase the difference between our estimate of the long-run effect and the estimate obtained from the conventional IV regression.

Our initial analysis of the data revealed a decline in persistence estimates in the post-1960 period. To be conservative when measuring the persistence of institutions, therefore, we estimate equation (14) using Constraints on the Executive data for the period 1900–1960s (presented in the baseline analysis) and for the period 1900–1990s (presented in the appendix).

Table 1 presents the results from our analysis.[18] Column 1 examines the case without any control variables. We estimate that raising the Constraints on the Executive in 1800 by one point on the 7-point index increases contemporary income per capita by 0.41 log points. This implies that increasing constraints from the lowest possible score (1) to the highest possible score (7) increases 1990 income per capita by approximately 2.3 standard deviations. While this is an economically significant effect, the estimated long-run coefficient is 37 percent smaller than the conventional IV estimate. Thus, accounting for the persistence in the endogenous explanatory variable is quantitatively important.

---

[15]Following recommendations by Albouy (2012) and Acemoglu et al. (2012), we use the log of potential settler mortality capped at 250 per 1000 as the instrument in the GMM regressions. The uncapped settler mortality variable is obtained directly from AJR (2001).

[16]The latitude variable is the latitude of a country's approximate geodesic centroid obtained from CIA's World Factbook. The regional dummies indicate the Sub-Saharan Africa, Middle East & North Africa, South Asia, East Asia and Pacific, and the North America regions, as defined by the World Bank. There are no observations from the Europe & Central Asia region and the Latin America & Caribbean region is the background region.

[17]Several studies have suggested that settler mortality is correlated with other contemporary variables, such as education or trade (e.g., Dollar and Kraay, 2003; Glaeser et al., 2004). For our results to be valid, we need only assume that settler mortality affected these other variables through historical institutions. Using the notation from Section 2.1, education or trade could serve as the A variable in our framework.

[18]Table A.1 in the appendix establishes that the results are robust to the use of data for the period 1900–1990s in the estimation of the persistence of institutions. In particular, the table establishes that the long-run estimates are even smaller when using data for that time period in the estimation of persistence. Furthermore, Tables A.2 and A.3 establishes that the results of Table 1 and Table A.1 are robust to the use of an alternative measure of contemporary income, i.e., GDP per capita in 2013.

Table 1: The Long-Run Effect of Institutions on Income Per Capita (1990s)

	Log GDP per capita in 1990s		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.411*	0.167	0.223
	(0.248)	(0.264)	(0.353)
Persistency of Endogenous Variable ($\hat{\delta}$)	0.835***	0.720*	0.791
	(0.216)	(0.432)	(0.483)
Conventional 2SLS Estimate	0.650***	0.384***	0.404***
	(0.136)	(0.106)	(0.102)
World Region Fixed Effects	No	Yes	Yes
Asolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.444	0.518	0.665
First Stage F-Statistic (K-P) of Conventional	30.968	5.755	5.740
First Stage F-Statistic (K-P) of Persistence	26.217	12.522	11.152
Number of Observations	56	56	56

This table presents the results of a series of augmented IV-regressions of the log average GDP per capita in the 1990s on an index of the historical level of constrains on the executive. It reports the estimated long-run effect of constraints on the executive in 1800 ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1960s ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of constraints on the executive. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

We find that the conventional IV regression overestimates the long-run effect. This occurs because institutions are less than perfectly persistent, i.e., $\delta < 1$. An increase in Constraints on the Executive in the contemporary period of one unit, therefore, corresponds to an increase in the 1800 measure of more than one unit. Unfortunately, the long-run estimate cannot be measured with much precision. The estimate is only significantly different from zero at the 10 percent level. This may reflect the fact that our estimate of δ is also not precisely estimated. The lack of precision is likely due to the small sample. In the next subsection, we use panel data to estimate δ with a larger sample. We again find that δ is significantly smaller than one.

Settler mortality may be correlated with many other geographic factors that affect income per capita, creating a classic violation of the exclusion restriction. Thus, the remainder of the table adds latitude and World Bank region fixed effects to the analysis.[19] The qualitative results are similar in all specifications. Column 3 presents our preferred specification. In this case, increasing Constraints on the Executive from the lowest possible score (1) to the highest possible score (7) increases 1990

[19]Acemoglu et al. (2001) use latitude and continent fixed effects as baseline controls. We use World Bank region fixed effects, which are more appropriate for a modern context and yield larger first stage F-statistics (Ashraf and Galor, 2013), indicating that the instrument is stronger when conditioning on these regional fixed effects compared to when conditioning on the continent fixed effects.

income per capita by approximately 1.27 standard deviations. This long-run coefficient is 45 percent smaller than the conventional IV estimate.

3.1.2 Assessment of Imperfect, Constant, and Linear Persistence

In the section, we use panel data to support our findings and validate key assumptions. We start by re-estimating the persistence of institutions. Then, we also use panel data to provide evidence for the assumptions that the persistence of institutions is linear and that δ is constant over time.

To measure the persistence of institutions using this panel data, we employ the panel-model analog of equation (11):

$$X_{c,t} = \alpha_c + \nu_t + \delta X_{c,t-1} + \varepsilon_{c,t}, \tag{18}$$

where X is a measure of institutional quality, ν_t is a time period fixed effect, α_c is a country fixed-effect. We use a superset of the Constraints on the Executive data from the Polity IV dataset that covers every year since 1850. When comparing equation equation (18) with equation (10) in Section 2.1, it can be seen that δ in equation (18) is indeed the relevant measure of institutional persistence. We run the regressions for different period lengths. In particular, we use yearly data, five-year data, and 10-year data. In the cases of five- and 10-year data, we average the data over each period.[20] Unlike the main analysis, we do not have a explicit source of variation in institutional quality, and the results may suffer from omitted variable bias. In this context, however, omitted variables are likely to affect past and current institutions in the same direction, biasing our estimate of δ upward.[21]

Imperfect Persistence

In our main analysis, we found $\delta < 1$, which implies that augmented estimate of the long-run effect is lower than the conventional IV coefficient. In this subsection, we provide alternate estimates of δ by running a series of panel regressions accounting for country-specific and period-specific fixed effects. We run the regressions for both the full dataset of 158 countries and for the sample of 56 countries included in the main analysis.

Table 2 presents the results. The point estimates suggest a low degree of persistence over the available period, which spans the years 1800–2013. This conclusion is robust to the sample used. Thus, our panel data analysis supports the finding that $\delta < 1$. Indeed, extrapolating the panel analysis to the 190-year time span indicates that the primary analysis may overestimate the persistence of the Constraints on the Executive. This would imply that it underestimates the quantitative impact of accounting for persistence. In other words, these results suggest that our main analysis provides a conservative correction of the conventional IV regression.

[20] Due to the existence of missing data in some period-country pairs, the averaging of observations increases the sample size. Furthermore, it increases the amount of data that is used for each period, relative to the inclusion of just one year of data for each period. The averaging, therefore, may help counter a possible attenuation bias. Note that, by averaging within periods, we are implicitly making the assumption that missing data in each period for each country is equal to the average of the non-missing data in each period for each country.

[21] For this reason, we do not include any time-varying controls. Without a more complete theory of institutional persistence, it is not possible to decide *a priori* which time-varying factors are channels of institutional persistence and are omitted variables.

Table 2: Panel Data Estimates of Persistence

	Constraint on the Executive					
	Main Sample (56 Countries)			Full Sample (158 Countries)		
	1-Year	5-Year	10-Year	1-Year	5-Year	10-Year
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Constraint on the Executive	0.916***	0.816***	0.696***	0.938***	0.848***	0.705***
	(0.013)	(0.024)	(0.034)	(0.005)	(0.011)	(0.015)
Number of Observations	11,928	2,352	1,176	54,954	10,836	5,418
Number of Countries	56	56	56	258	258	258
Adjusted R^2	0.910	0.821	0.725	0.906	0.785	0.638
Test of $\delta = 1$ (p-Value)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

This table presents the results of a series of panel regressions of an index of the historical level of constrains on the executive on its lagged values. The regressions account for country-specific and period-specific fixed effects. Furthermore, the table reports the results of a Wald test of the null hypothesis that the estimate of the persistence coefficient is equal to one. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

Furthermore, though our panel data results show surprisingly little persistence in the Constraints on the Executive, they are consistent with the existing literature. A growing literature examines the determinants of institutions, focusing on whether increases in income facilitates democratization (the 'Modernization Hypothesis'). While it is not the goal of these papers to measure institutional persistence, the lag of institutions is often included as a control. In this literature, the coefficient on lagged institutions is significantly less than one, providing further support for our results (Acemoglu et al., 2008, 2009; Heid et al., 2012; Benhabib et al., 2011; Cervellati et al., 2014).

Constant Persistence

We now investigate the assumption that δ is constant over time. To do so, we run rolling panel data regressions with a 50-year window. In particular, for each 50-year period starting in the years between 1850 and 1963, we run a regression based on equation (18).[22] We then plot the estimate and its 95 percent confidence interval against the initial year of each sample. The range and standard deviation of the coefficient estimates provide insight into the stability of δ over time.

[22] To compare estimates of δ over time, it is important that the sample of countries does not change substantially from period to period. We therefore restrict the sample to those countries that are in the main IV sample and for which data on executive constraints exist for at least 75 percent of the years in the period 1850–2013. This results in a sample of 21 countries. As established in Table A.1, the results are robust to the use of data on the 47 countries out of all countries in the Polity IV database for which data on Constraints on the Executive exists for at least 75 percent of the years in the period 1850–2013. Furthermore, as established in Table A.2, the results are also robust to the inclusion of all the 158 countries in the Polity IV database for which data on Constraints on the Executive exists for at some years in the period 1850–2013.

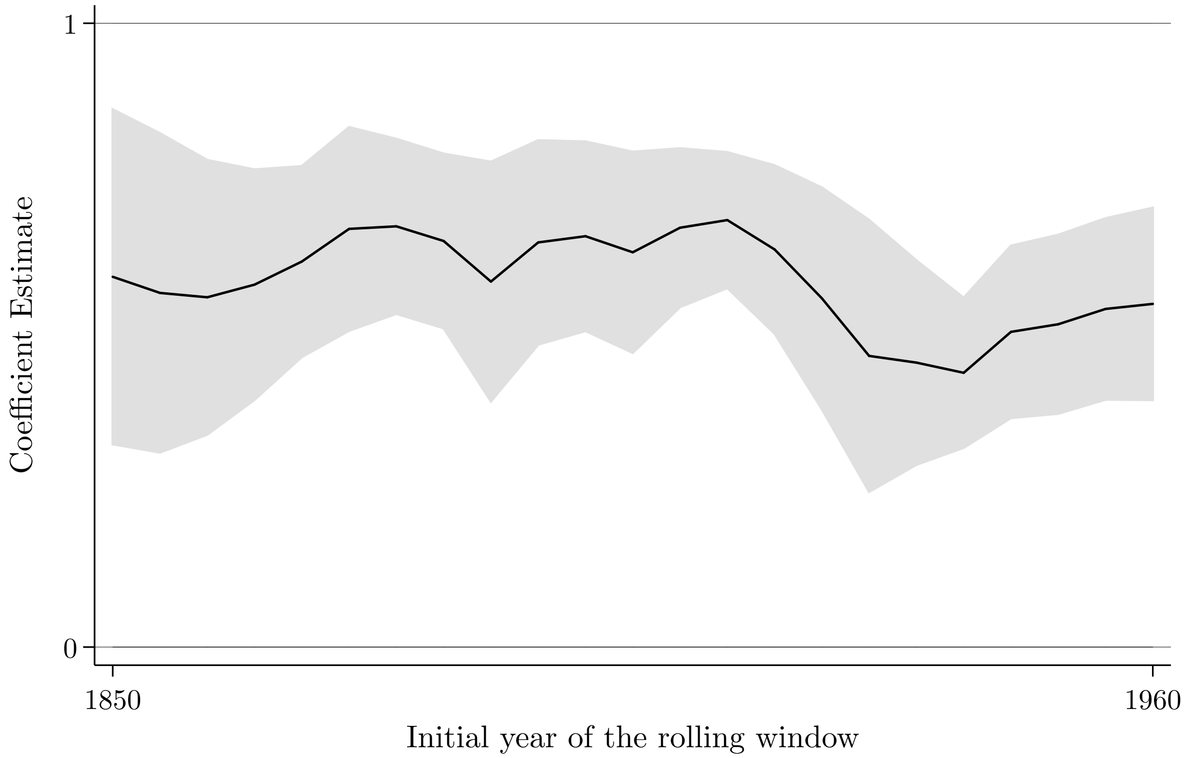

Figure 2: This figure depicts the coefficient from five-year panel regressions of Constraints on the Executive on its lagged value in over period 1850–2013 with a 50-year regression window and a step size of five years, estimated with OLS. The sample is restricted to those 21 countries, out of the sample of 56 countries from the main analysis, for which information on Constraints on the Executive exists in the Polity IV database for at least 75 percent of the years in the period 1850–2013. The regressions account for country and year fixed effects. Robust standard errors are used for the calculation of the confidence band.

Furthermore, the plotted coefficient estimates provides insight into possible trends in the estimates of δ.

The results are presented in Figure 2. There are two main takeaways from this analysis. First, the coefficient on lagged institutions appears relatively stable, hovering between 0.43 and 0.68, with a mean of 0.56. The standard deviation of the coefficients is just 0.07. Moreover, the figure does not reveal any obvious time trends in the estimate of δ. This stability of the estimated persistence coefficient suggests that our assumption of a constant δ is a reasonable approximation. Second, the estimate is always significantly below one, which reinforces our finding of a long-run effect that is smaller than the conventional 2SLS coefficient estimate.

Linear Persistence

Finally, we use the panel dataset to examine whether the persistence of Constraints on the Executive is linear. We do so by examining the non-parametric fit of the relationship between the variable and its lagged value, partialling out the country and period specific fixed effects. Comparing the results to a linear fit allows us to test whether our assumption is a reasonable approximation.

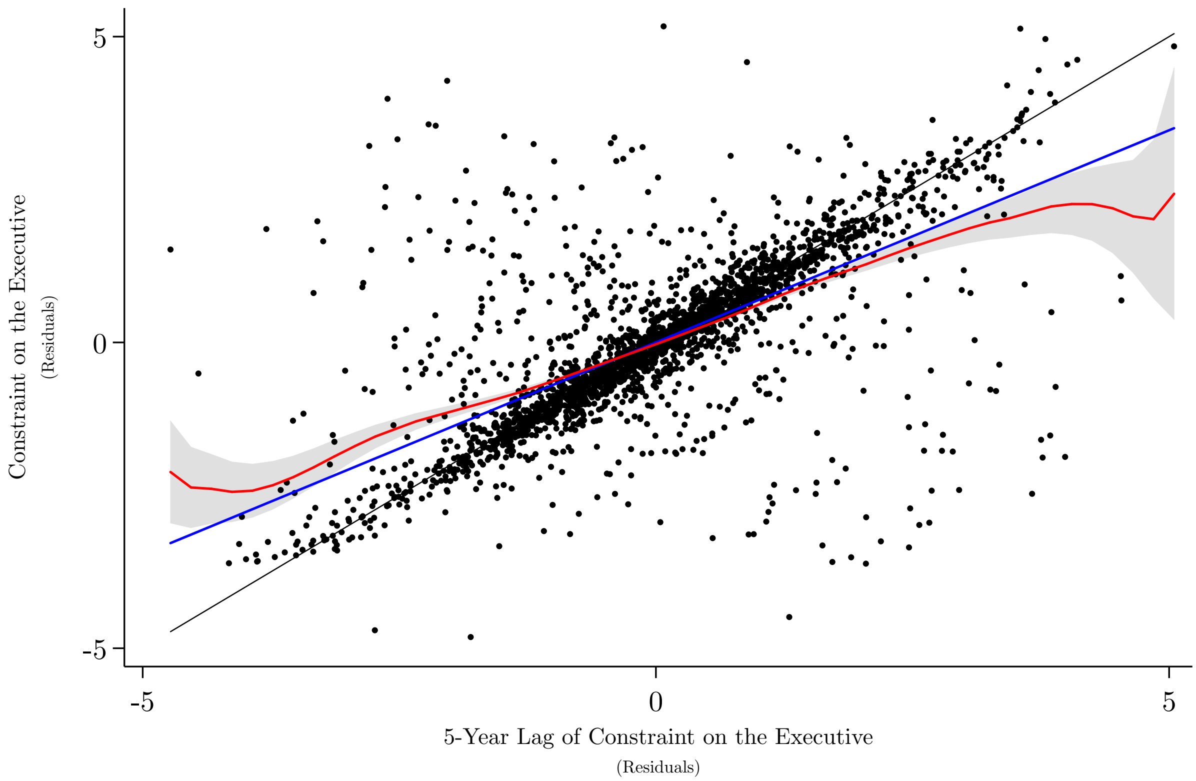

Figure 3: Linear and flexible fits of δ using a 5-year lagged panel data model accounting for country and year fixed effects. The black line represents $\delta = 1$. The blue line represents the fit from a linear regression model using a linear specification. The red line represents a flexible fit from a kernel-weighted local-mean smoothing. The shaded area represents the 95 percent confidence bounds of the flexible fit.

To estimate the relationship between Constraints on the Executive and its lagged value non-parametrically, we first run regressions of X_t and X_{t-1} on the time and country fixed effects using a period length of five years data.[23] We then capture the residuals from each regression. We then run linear regression of the residuals from the current period regression on the residuals from the lagged regression. The slope of the linear fit is, by construction, equal to δ from the equation (18), but we can now visually examine the relationship. Moreover, we use the two sets of results to construct a flexible estimate of δ using kernel-weighted local-mean smoothing.[24]

The results are presented in Figure 3. Importantly, the non-parametric and linear regression lines are generally very close to one another, and the linear fit is typically within the 95 percent confidence band around the non-parametric fit. The non-parametric fit only deviates somewhat notably from the linear fit in the sparse extremes of the Constraints on the Executive index.[25] The

[23]The conclusion of the linearity assessment is robust to the use of alternative lag lengths. In particular, Table A.3 in the appendix establishes that the non-parametric fit remains approximately linear when using a 10-year data.

[24]We use an Epanechnikov kernel and a rule-of-thumb bandwith as defined in Stata 14's lpoly command.

[25]Furthermore, fitting a linear regression with a quadratic specification reveals that the second-order term is very close to zero (.003) and insignificant ($p = .685$), again indicating that a linear fit is appropriate.

similarity between the linear and the non-parametric fit suggests that linearity in past levels of Constraints on the Executive is a valid assumption.[26]

Overall, these findings provide support to the conclusions reached in our main analysis in Section 3.1.1. We confirm the key finding that $\delta < 1$ in panel data and provide evidence to support our two main assumptions. Thus, we have strong reason to believe that long-run effect is smaller than the conventional 2SLS estimate, demonstrating the importance of accounting for the persistence of the endogenous variable when estimating the long-run effect.

3.2 Religion and Literacy

In our second application, we build on the influential work of Becker and Woessmann (2009) and examine the effect of the spread of Protestantism on human capital accumulation in Prussia between 1517 and 1870. Becker and Woessmann (2009) revisit the work of Weber (1958), but stress human capital accumulation, rather than work ethic, as the mediating factor between Protestantism and positive economic outcomes. In particular, they examine the causal effect of the Protestant share of a Prussian county in 1870 on literacy and distance to the near school, also in 1870. The instrument is the distance from Wittgenstein, the location where Martin Luther penned, and reportedly hung, the 95 theses in 1517. The underlying theory is that Protestantism spread out geographically from Wittgenstein, implying that closer counties should have higher Protestant shares. Thus, the shock from the instrument occurred in 1517 and the endogenous regressor is measured in 1870.[27] The considerable time gap between the initial shock and the measurement of the endogenous variable implies that our framework is relevant in this context.

There are several possible A variables in this setting. First, the historical (1500s) protestant share could affect the historical level of income – consistent with the author's finding of an association between the protestant share and income in 1870 – which could then affect literacy in 1870. Alternatively, there could be persistence in human capital accumulation. This could occur, for example, because increases in education increase the supply of teachers and because schools lower the cost of acquiring education. Finally, the A variable could capture feedback between human capital and fertility. Using the same data and setting, Becker et al. (2010) demonstrate that increases in education during this period led to decreases in fertility, and that decreases in fertility led to increases in education. Thus, fertility would be the A variable in our framework. Table 8 in the appendix provides summary statistics.

It is possible to apply our method to this example because of the availability of data on the endogenous explanatory variable for more than one period in time. The *ifo Prussian Economic History Database (iPEHD)* includes the variables used in the original study as well as the protestant

[26]We also examine the non-linearity of the relationship using the data from the main analysis with OLS regressions (please see Figure A.4 in the appendix). We again find that the relationship development in Constraints on the Executive is approximately linear.

[27]The exact timing of the shock is fuzzy, which further complicates the interpretation of regressions of this type. We think 1517 is a useful approximation to the time period of the initial shock.

Table 3: The Long-Run Effect of Protestantism on Literacy

	Percentage Literate in 1871		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1571}$)	0.196***	0.172***	0.178***
	(0.034)	(0.016)	(0.017)
Persistency of Endogenous Variable ($\hat{\delta}$)	0.983***	0.988***	0.985***
	(0.007)	(0.009)	(0.009)
Conventional 2SLS Estimate	0.215***	0.183***	0.193***
	(0.036)	(0.014)	(0.015)
Control Variable Group 1[a]	No	Yes	Yes
Control Variable Group 2[b]	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.018	0.206	0.087
First Stage F-Statistic (K-P) of Conventional	75.821	428.101	378.459
First Stage F-Statistic (K-P) of Persistence	74.478	392.718	349.576
Number of Observations	280	280	280

This table presents the results of a series of augmented IV-regressions of the percentage of literate individuals in 1871 on the historical percentage of Protestants. It reports the estimated long-run effect of the percentage of Protestants in 1517 ($\hat{\eta}_{1517}$), the estimated persistence of the percentage of Protestants from 1816 to 1871 ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of the percentage of Protestants. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

[a]Longitude (in radians), latitude (in radians), and longitude × latitude (in radians).

[b]University in 1517, imperial city in 1517, Hanseatic city in 1517, monestaries in 1517 per square kilometer, school in 1517, and city population in 1500.

share in 1817 (Becker et al., 2014).[28] Thus, we can use the persistence in the Protestant share between 1817 and 1870 to estimate δ and apply our augmented estimator. We focus on literacy as our main dependent variable, but also present results for distance to the nearest school in the appendix. Since the measure of persistence is the same in both cases, the quantitative impact of applying the augmented estimator is also the same. We take all control variables from the original study, but our main specification is different. In particular, we do not want to include endogenous controls that could serve as part of the A variable, because this would complicate the interpretation of the results. Thus, we only include purely exogenous control variables used in the original study.

Table 3 presents the results. Column 1 presents results without any control variables. The magnitude of our estimate of η suggests that increasing the Protestant share in a given county in 1517 from 0 percent to 100 percent raises 1870 literacy by 1.7 standard deviations.[29] This

[28]We follow the nine-step procedure for merging information intertemporally in the iPEHD database at `https://www.cesifo-group.de/ifoHome/facts/iPEHD-Ifo-Prussian-Economic-History-Database/merging.html`. For all variables representing proportions of the population, we calculate the population-weigthed averages across observations sharing the same locational identifier. For all other variables, we calculate the unweighted averages.

[29]See Table A.8 for summary statistics.

effect is economically and statistically significant. It is 9 percent lower than the estimate from the conventional IV regressions, which again indicates the quantitative importance of our new method. When compared to the previous applications, the estimates are more precise, likely due to the larger sample. In particular, we can reject the null hypothesis that $\delta = 1$ at the 5 percent level. We can also reject the null hypothesis that $\eta = 0$ at the 1 percent level.

The remainder of the table repeats this exercise when adding control variables. In column 2, we add longitude, latitude, and their interaction to capture geographic factors that may be correlated with distance to Wittgenstein. In column 3, we add initial county-level characteristics, including the type of city, number of schools, number of monasteries, and the percentage of the population living in a city. In all cases, the results are qualitatively and quantitatively similar. In particular, our preferred specification in column 3 yields an estimate of the long-run effect that is 8 percent smaller than the conventional IV estimate.

Table A.4 in the appendix is analogous to the main table, but uses distance to the nearest school, rather than literacy, as the dependent variable. As demonstrated in equation (17), the difference between η and the conventional regression coefficient is fully determined by the persistence. Thus, the ratios between the conventional and augmented estimates in this table is equal to the those implied by Table 3.

Tables A.5 and A.6 show the results when including the endogenous controls from the original study using literacy and distance to school as outcome variables, respectively. In some specifications, we find persistence that is greater than one, but it is difficult to interpret these results given that the endogenous controls likely capture part of A and may introduce other biases.

Unfortunately, we do not have panel data on protestant share to test our assumptions of constant and linear persistence. Figure A.5 in the appendix, however, presents a cross-section plot comparing the extent of Protestantism in 1817 and 1870. The relationship is highly linear, providing some support for this assumption.

4 Discussion

In this section, we provide a broader discussion of our findings. In Subsection 4.1, we explain how applied researchers can apply the insights and results from this paper. In Subsection 4.2, we discuss how our framework can help illuminate the interpretation of related empirical techniques.

4.1 Practical Steps

Thus far, we have discussed how to interpret regressions where the instrument precedes the endogenous regressor in time and provided a method for estimating the long-run effect of changes in historical conditions. In this subsection, we provide a summary of our findings while paying special attention to their implications for applied research.[30]

The first step when investigating regressions where the instrument precedes the endogenous regressor is time is to determine which effect, the contemporaneous effect (β_1) or the long-run

[30] A Stata program for application of our method is available upon request and will be released with the paper.

effect (η), is the parameter of interest. If the goal is to estimate the contemporaneous effect, then the standard estimators would estimate the parameter of interest if and only if there is no A variable. If an A variable does exist, then it is not possible to estimate the desired parameter, even after applying the new method developed here. If the goal is to estimate the long-run effect, then it is necessary to account for the persistence of the endogenous variable, whether or not an influential A variable exists.

The next step, therefore, is to consider theoretically whether an A variable exists or not. As argued above, we believe that an A will exist in most applications. In particular, whatever process links contemporary values of the endogenous variable, X_C, to the outcome of interest, Y, will also generally link past values of the endogenous variable, X_H, to past values of the outcome variable. If this process has a persistent component — or if the outcome variable has a persistent component — then there will be a link between X_H and Y that does not operate through X_C.

Finally, if the researcher is interested in evaluating the long-run effect of the endogenous variable, it is necessary to consider the degree of persistence, δ. If there is no strong theoretical reason for the endogenous variable to be perfectly persistent — i.e., $\delta = 1$ — the researcher should investigate if the assumptions of our framework are applicable. The key assumptions are (i) that it is possible to observe persistence of the endogenous variable at two intermediate points in time and (ii) that the intermediate data are suitable to generate a reasonable extrapolation of the persistence over the entire time period. Assumption (i) is a matter of data availability. Section 3.1.2 demonstrates how to investigate assumption (ii) using panel data in the case that persistence is assumed to be constant and linear. This framework could be extended to alternative functional forms and a time-varying δ. If the researcher determines that it is not possible to use observed data to infer the long-run persistence, then it is not possible to directly estimate the long-run effect. In the presence of an A variable, this implies that neither β_1 nor η can be estimated with the conventional regressions. It may still be possible to bound the estimate of the long-run effect, if the researcher has a strong prior regarding the bounds of δ.

4.2 Perspectives on Related Empirical Techniques

Thus far, we have focused on instrumental variable regressions where the instrument precedes the endogenous regressor in time. Our framework, however, can provide insight into related empirical techniques. In this subsection, we provide a quick discussion of some of these alternate methodologies.

4.2.1 Identification with Gravity-based Instruments

Our main analysis focuses on cases where the instrument only affects contemporary values of the endogenous variable via historical values. A closely related situation occurs when the instrument might directly affect both historical and contemporary values of the endogenous explanatory variable. The most prominent case falling into this category would be cross-sectional gravity-based regressions measuring the impact of trade or migration on income per capita (Frankel and Romer,

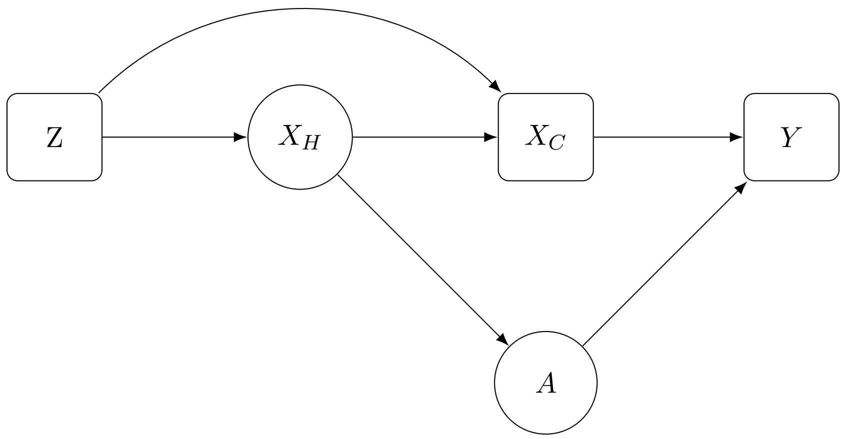

Figure 4: Causal diagram for gravity-based instruments. Rectangular nodes represent observed variables and circular nodes represent unobserved variables. Unlike the basic framework presented in Figure 1, there is a direct causal effect from the historical instrument, Z, the to the contemporary value of the endogenous regressor, X_C. In this case, neither coefficient of interest is recoverable with the conventional regression or our augmented approach.

1999; Do et al., 2016; Alesina et al., 2016). In these applications, researchers use time-invariant geographical characteristics to construct an exogenous instrument for a country's propensity to trade. In this case, geography affects both contemporary and past values of the endogenous variable, and it is less clear whether the inherent parameter of interest is β_1 or η.

Figure 4 extends our basic framework to the case where Z affect X_C directly, as in the case of the gravity-based trade regressions. As the figure readily highlights, Z would no longer be a valid instrument for X_H, even if the latter could be observed. In this case, neither η nor β_1 can be recovered from the regression, even after employing our new method. Thus, while we cannot provide a remedy, our analysis highlights the difficultly in interpreting results generated by this prominent regression technique.[31]

4.2.2 Regression Models with Ancestry-adjusted Variables

In the study of long-run economic growth, it is often important to distinguish between the characteristics of a physical location and the characteristics of people living in that location. Recent work by Putterman and Weil (2010) allows researchers to adjust historical determinants of development by population flows in the post-1500 period. Thus, for any historical variable, it is possible to construct a measure that is relevant for the people living in a location, even when their ancestors may have lived elsewhere. Though the underlying technique is OLS, many of the insights discussed here are relevant for this literature.

In the context of our framework, we can think of X_H as the unadjusted variables, X_C as the adjusted variable, and Y as the contemporary outcome variable.[32] Then, δ captures the effect

[31] An approach developed by Feyrer (2009a,b) uses time-varying geographical instruments, which estimates the contemporary effect of trade on income (β_1) without being affected by the issues raised in our paper.

[32] See Appendix Section A.5 for a formal analysis of this case.

of the unadjusted variable on the adjusted variable. If international migration is not orthogonal to X_H, then δ will differ from one. For example, if countries with good historical institutions attract population flows from other countries, this will affect the adjusted level, X_C. Similarly, countries with poor historical institutions might be vulnerable to colonization, which would also affect X_C. As long as there are no other threats to identification, a regression of Y on X_C will yield a regression coefficient similar to that given in equation (5). In other words, the resulting regression coefficient will be the long-run effect of the unadjusted variable divided by the persistence between the unadjusted and adjusted variable.

Luckily, in this case, there is a simple solution to identifying β_1, which is the coefficient of interest in this context.[33] In particular, β_1 can be estimated by simultaneously including the unadjusted and adjusted variables in the regression. In this case, the coefficient on the adjusted measure will estimate β_1, while the coefficient on the unadjusted measure will capture $\gamma\beta_2$, the effect of X_H on Y via A. Some prominent papers in the literature already perform such regressions to compare effects of the unadjusted and adjusted variable (e.g., Putterman and Weil, 2010; Ashraf and Galor, 2013). Thus, our results give new interpretation to these regressions and suggest that they should be the main specification moving forward. Some of these analyses even suggest that the effect of adjusted and unadjusted variables might have opposite signs, further highlighting the need to separate the effects of each component (e.g., Comin et al., 2010; Galor and Özak, 2016).

4.2.3 Historical Endogenous Regressors

The issues of interpretation and identification that arise in this paper occur because historical values of the endogenous regressor, X_H, cannot be observed. Thus, the concerns raised here can be overcome with the collection of historical data, which is a promising trend in the field of long-run economic growth (e.g., Nunn, 2008; Iyer, 2010; Bruhn and Gallego, 2012). To take a simple example, the work of Banerjee and Iyer (2005) and Feyrer and Sacerdote (2009) address the question of how institutions affect economic development with historical data on institutional quality, directly estimating η. Our results suggest that further integration of long-run growth and archival-based historical research can overcome identification problems in existing techniques.

4.2.4 Spatial Regression Discontinuity

Spatial regression discontinuity presents an exciting way to achieve identification without being subjected to the issues discussed here (e.g., Dell, 2010; Michalopoulos and Papaioannou, 2013). Using the lens of our framework, the benefit of spatial regression discontinuity is that the historical shock can be perfectly observed in the contemporary period. In particular, we can think of Z as being a random shock, X_H and X_C as being an indicator variables that take a value of one at certain geographic locations, and Y as contemporary levels of development. In this case, δ is exactly equal to one in all cases, and observations of X_C (contemporary location on the earth) are sufficient to measure X_H (historical location on the earth). Thus, the regression coefficient will

[33] It possible to estimate η by regressing the outcome variable on the unadjusted measure.

accurately capture the long-run effect, η, of a historical shock through all channels. In this case, a particularly interesting aspect of an A variable will be spatial sorting. After the initial shock, individuals will sort across the regions, mitigating the differences in the outcome variable. Thus, a key component to understanding the channels through which historical shocks affect contemporary development would be to understand the limits to sorting.

5 Conclusion

A growing literature convincingly argues that historical events continue to shape contemporary economic development (Spolaore and Wacziarg, 2013; Nunn, 2014; Ashraf and Galor, 2016). Often, however, we do not know how to interpret the magnitude and economic importance of these findings. This hinders the translation of the findings into policy advice that is relevant for developing countries. We provide a framework to understand a popular methodology, namely IV regressions where the instrument precedes the endogenous regressor in time, and investigate the interpretation of the regression coefficients. We then provide an augmented estimator that estimates the long-run effect of changes in historical conditions. We then apply our results to two prominent topics in the field of long-run economic growth. We also discuss how our findings related to other empirical techniques used in the study of long-run economic growth.

Economics is making exciting progress towards understanding the complex process of long-run economic development, both empirically and theoretically. A key implication of our work is that these two approaches cannot be fully separated. To understand the implications of econometric work, it is necessary to consider the theory informing the underlying data generating process. This paper demonstrates that even a very simple formal presentation of long-run dynamics can greatly alter our understanding of the interpretation and limitations of commonly used econometric techniques. In this way, our results are closely related to works by Acemoglu (2010) and Deaton (2010a,b), who stress the importance of utilizing theory to make sense of empirical results in economic development. Recent work by Cervellati and Sunde (2015) and Andersen et al. (2016) explicitly consider the relationship between long-run dynamics and empirical results in the field of economic growth. In light of our analysis, this type of work presents an exciting way forward to better understand the mechanisms of economic development.

A Further Algebraic Implications

A.1 No Alternative Channels

In this subsection, we examine the interpretation of the standard IV regression without the presence of an A variable. This is a special case of our more general framework. The simplified system is given by:

$$X_{H,i} = \kappa_{X_H} + \psi Z_i + \varepsilon_{X_H,i} \tag{19}$$

$$X_{C,i} = \kappa_{X_C} + \delta X_{C,i} + \varepsilon_{X_C,i} \tag{20}$$

$$Y_i = \kappa_Y + \beta X_{C,i} + \varepsilon_{Y,i}. \tag{21}$$

In this set-up, $\eta = \frac{\partial Y}{\partial X_H} = \delta\beta$. Since there is no violation of the exclusion restriction, the IV regression yields:

$$\text{plim}\,\hat{\beta}^{IV} = \beta. \tag{22}$$

Thus, as in the more general framework, $\eta = \delta\,\text{plim}\,\hat{\beta}_1^{IV}$. So, our results for estimating η hold in this special case. A key aspect of our paper is that this simple result still holds under certain violations of the exclusion restriction, specifically those that take the form of the A variable in Section 2. We believe that this is the empirically relevant case when investigating causes of long-run economic growth.

A.2 Reverse Causality and Historical Instruments

In this subsection, we discuss the role of reverse causality in interpreting the estimates of IV regressions in long-run economics growth. Our parameter of interest, η, actually incorporates reverse causality. To see this, we can add reverse causality to the framework of Section 2:

$$X_{H,i} = \kappa_{X_H} + \psi Z_i + \varepsilon_{X_H,i} \tag{23}$$

$$Y_{H,i} = \kappa_{Y_H} + \beta X_{H,i} + \varepsilon_{Y_H,i} \tag{24}$$

$$X_{C,i} = \kappa_{X_C} + \delta X_{H,i} + \varphi Y_{H,i} + \varepsilon_{X_C,i} \tag{25}$$

$$A_i = \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i} \tag{26}$$

$$Y_{C.i} = \kappa_{Y_C} + \beta_1 X_{C,i} + \beta_2 A_i + \varepsilon_{Y_C,i}. \tag{27}$$

Now, defining $\tilde{\eta} \equiv \frac{\partial Y_C}{X_H} = \beta_1\delta + \beta_2\gamma + \beta_1^2\varphi$ and $\tilde{\delta} \equiv \frac{\partial X_C}{\partial X_H} = \delta + \beta_1\varphi$, we find that the IV estimator yields:

$$\text{plim}\,\hat{\beta}_1^{IV} = \frac{\tilde{\eta}}{\tilde{\delta}}. \tag{28}$$

It is apparent that the reverse causality coefficient enters the expression for $\tilde{\eta}$. This doesn't change the fact that a one-unit change in X_H will increase Y_C by $\tilde{\eta}$, but it is necessary to keep in mind the limited ability of the IV estimator to help identify the effect of one isolated factor, even if historical data on X in employed.

A.3 Omitted Variables

We now consider what the IV regression accomplishes when compared to OLS. So far, we haven't introduced any explicit violations of the exclusion restriction other than causal channels, represented by A. Now, we consider the effect of a variable, W, that is correlated with X_H, but is not causally affected by X_H. Geographic characteristics are likely candidates for this type of variable. The system is now given by:

$$\begin{aligned} X_{H,i} &= \kappa_{X_H} + \psi Z_i + \varepsilon_{X_H,i} && (29)\\ X_{C,i} &= \kappa_{X_C} + \delta X_{H,i} + \varepsilon_{X_C,i} && (30)\\ A_i &= \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i} && (31)\\ Y_i &= \kappa_Y + \beta_1 X_{C,i} + \beta_2 A_i + \beta_3 W_i + \varepsilon_{Y,i}, && (32) \end{aligned}$$

where $Cov(W, X_0) \equiv \nu \neq 0$ but $Cov(W, Z) = 0$. We also define $\mathrm{Var}(X_0) \equiv \xi \neq 0$. In this case, the OLS coefficient picks up the association between W and X_H in the usual omitted variable fashion, but IV does not:

$$\begin{aligned} \text{plim}\, \hat{\beta}_1^{OLS} &= \beta_1 + \frac{\gamma\beta_2}{\delta} + \frac{\beta_3\nu}{\delta\xi} = \frac{\eta}{\delta} + \frac{\beta_3\nu}{\delta\xi} && (33)\\ \text{plim}\, \hat{\beta}_1^{IV} &= \beta_1 + \frac{\beta_2\gamma}{\delta} = \frac{\eta}{\delta}. && (34) \end{aligned}$$

Indeed, the IV coefficient is the same as in Section 2.1. So, the IV coefficient eliminates bias caused by correlates of X_H, but does not eliminate the effect of channels through which X_H directly affects Y.

A.4 Channels

Finally, we demonstrate that, if X_H affects X_C through a channel, researchers should not control for this channel when estimating the long-run effect. As demonstrated in the previous subsection, this distinguishes a channel from an omitted variable. Consider the following extensions of the results from Section 2:

$$\begin{aligned} X_{H,i} &= \kappa_{X_H} + \psi Z_i + \varepsilon_{X_H,i} && (35)\\ A_i &= \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i} && (36)\\ U_i &= \kappa_U + \tau X_{H,i} + \varepsilon_{U,i} && (37)\\ X_{C,i} &= \kappa_{X_C} + \delta X_{H,i} + \xi U_i + \varepsilon_{X_C,i} && (38)\\ Y_i &= \kappa_Y + \beta_1 X_{C,i} + \beta_2 A_i + \varepsilon_{Y,i}. && (39) \end{aligned}$$

Plugging (36) into (38) yields:

$$X_{C,i} = \kappa_{X_C} + \xi\kappa_U + (\delta + \xi\tau) X_{H,i} + \varepsilon_{X_C,i} + \xi\varepsilon_{U,i}. \tag{40}$$

Now, defining $\tilde{\delta}_1 \equiv \delta + \xi\tau$, we have a system that is similar to that in Section 2, except that $\tilde{\delta}$ defines the persistence of institutions. Thus, to estimate the long-run effect via the method presented in this paper, it is necessary to estimate the total persistence, $\tilde{\delta}$.

A.5 Ancestry-Adjustment

In this section, we derive the results discussed in Section 4.2.2. Start with the basic algebraic framework:

$$\begin{aligned} X_{C,i} &= \kappa_{X_C} + \delta X_{H,i} + \varepsilon_{X_C,i} \quad (41) \\ A_i &= \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i} \quad (42) \\ Y_i &= \kappa_{Y_H} + \beta_1 X_{C,i} + \beta_2 A_i + \varepsilon_{Y,i}, \quad (43) \end{aligned}$$

where $X_{H,i}$ is the unadjusted measure and $X_{C,i}$ is the adjusted measure. Substituting (42) into (43) yields

$$Y_i = \kappa_{Y_H} + \beta_2 \kappa_{A_H} + \beta_1 X_{C,i} + \beta_2 \gamma X_{H,i} + \varepsilon_{Y,i} + \beta_2 \varepsilon_{A,i}. \quad (44)$$

As long as both measures of X are orthogonal to the error term, estimating equation (44) with an OLS regression of the outcome variable on both the adjusted and unadjusted measures will yield the coefficients discussed in the text. On the other hand, if only the adjusted measure is included, the regression will yield,

$$\text{plim}\, \hat{\beta}_1^{AA} = \beta_1 + \frac{\beta_2 \gamma}{\delta} = \frac{\eta}{\delta}, \quad (45)$$

which is exactly the result derived from the IV analysis. Thus, the interpretation of the regression is subject to the same concerns discussed throughout this paper.

B Additional Tables

Table A.1: The Long-Run Effect of Institutions on Income Per Capita (1990s) — Using the Period 1900-1990s for the Estimation of Persistence

	Log GDP per capita in 1990s		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.407**	0.146	0.154
	(0.175)	(0.178)	(0.191)
Persistency of Endogenous Variable ($\hat{\delta}$)	0.761***	0.545*	0.500*
	(0.170)	(0.301)	(0.285)
Conventional 2SLS Estimate	0.712***	0.508**	0.639***
	(0.169)	(0.199)	(0.197)
World Region Fixed Effects	No	Yes	Yes
Asolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.159	0.130	0.079
First Stage F-Statistic (K-P) of Conventional	27.064	5.292	5.349
First Stage F-Statistic (K-P) of Persistence	26.217	12.522	11.152
Number of Observations	56	56	56

This table presents the results of a series of augmented IV-regressions of the log average GDP per capita in the 1990s on an index of the historical level of constrains on the executive. It reports the estimated long-run effect of constraints on the executive in 1800 ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1990s ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of constraints on the executive. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

Table A.2: The Long-Run Effect of Institutions on Income Per Capita (2013) — Using the Period 1900-1960s for the Estimation of Persistence

	Log GDP per capita in 2013		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.442*	0.167	0.224
	(0.265)	(0.248)	(0.334)
Persistency of Endogenous Variable ($\hat{\delta}$)	0.835***	0.720*	0.791
	(0.216)	(0.432)	(0.483)
Conventional 2SLS Estimate	0.700***	0.385***	0.406***
	(0.145)	(0.096)	(0.099)
World Region Fixed Effects	No	Yes	Yes
Asolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.444	0.518	0.665
First Stage F-Statistic (K-P) of Conventional	30.968	5.755	5.740
First Stage F-Statistic (K-P) of Persistence	26.217	12.522	11.152
Number of Observations	56	56	56

This table presents the results of a series of augmented IV-regressions of the log average GDP per capita in 2013 on an index of the historical level of constrains on the executive. It reports the estimated long-run effect of constraints on the executive in 1800 ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1960s ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of constraints on the executive. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

Table A.3: The Long-Run Effect of Institutions on Income Per Capita (2013) — Using the Period 1900-1990s for the Estimation of Persistence

	Log GDP per capita in 2013		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.438**	0.146	0.155
	(0.186)	(0.164)	(0.178)
Persistency of Endogenous Variable ($\hat{\delta}$)	0.761***	0.545*	0.500*
	(0.170)	(0.301)	(0.285)
Conventional 2SLS Estimate	0.768***	0.509***	0.642***
	(0.181)	(0.174)	(0.177)
World Region Fixed Effects	No	Yes	Yes
Asolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.159	0.130	0.079
First Stage F-Statistic (K-P) of Conventional	27.064	5.292	5.349
First Stage F-Statistic (K-P) of Persistence	26.217	12.522	11.152
Number of Observations	56	56	56

This table presents the results of a series of augmented IV-regressions of the log average GDP per capita in 2013 on an index of the historical level of constrains on the executive. It reports the estimated long-run effect of constraints on the executive in 1800 ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1990s ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of constraints on the executive. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

Table A.4: The Long-Run Effect of Protestantism on Distance to School

	Distance to School > 3 km		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1571}$)	-0.038***	-0.052***	-0.050***
	(0.009)	(0.008)	(0.008)
Persistency of Endogenous Variable ($\hat{\delta}$)	0.983***	0.988***	0.985***
	(0.007)	(0.009)	(0.009)
Conventional 2SLS Estimate	-0.042***	-0.055***	-0.055***
	(0.010)	(0.008)	(0.008)
Control Variable Group 1[a]	No	Yes	Yes
Control Variable Group 2[b]	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.018	0.206	0.087
First Stage F-Statistic (K-P) of Conventional	75.821	428.101	378.459
First Stage F-Statistic (K-P) of Persistence	74.478	392.718	349.576
Number of Observations	280	280	280

This table presents the results of a series of augmented IV-regressions of a dummy variable indicating if the distance to school in 1871 is above 3 km on the historical percentage of Protestants. It reports the estimated long-run effect of the percentage of Protestants in 1517 ($\hat{\eta}_{1517}$), the estimated persistence of the percentage of Protestants from 1816 to 1871 ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of the percentage of Protestants. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

[a]Longitude (in radians), latitude (in radians), and longitude $\times$ latitude (in radians).

[b]University in 1517, imperial city in 1517, Hanseatic city in 1517, monestaries in 1517 per square kilometer, school in 1517, and city population in 1500.

Table A.5: The Long-Run Effect of Protestantism on Literacy

	Percentage Literate in 1871							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Long-Run Effect ($\hat{\eta}_{1571}$)	0.122***	0.165***	0.159***	0.218**	0.147***	0.184***	0.111***	0.107***
	(0.026)	(0.034)	(0.040)	(0.108)	(0.034)	(0.038)	(0.025)	(0.025)
Persistency of Endogenous Variable ($\hat{\delta}$)	1.005***	1.021***	1.032***	0.971***	1.013***	1.012***	1.000***	0.991***
	(0.011)	(0.016)	(0.024)	(0.051)	(0.014)	(0.013)	(0.011)	(0.007)
Conventional 2SLS Estimate	0.119***	0.147***	0.134***	0.255**	0.137***	0.172***	0.110***	0.112***
	(0.024)	(0.026)	(0.032)	(0.101)	(0.028)	(0.030)	(0.023)	(0.024)
Control Variable Group 1[a]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variable Group 2[b]	No	Yes	Yes	Yes	No	No	No	No
Control Variable Group 3[c]	No	No	Yes	Yes	No	No	No	No
Control Variable Group 4[d]	No	No	No	Yes	No	No	No	No
Control Variable Group 5[e]	No	No	No	No	Yes	No	No	No
Control Variable Group 6[f]	No	No	No	No	No	Yes	No	No
Wald Test of $\hat{\delta} = 1$ p-value	0.692	0.195	0.185	0.572	0.329	0.349	0.966	0.220
First Stage F-Statistic (K-P) of Conventional	92.636	57.717	70.352	6.037	86.457	64.406	86.463	75.641
First Stage F-Statistic (K-P) of Persistence	84.758	51.550	58.377	5.840	78.009	58.308	80.037	71.258
Number of Observations	280	280	280	280	280	280	249	219

This table presents the results of a series of augmented IV-regressions of the percentage of literate individuals in 1871 on the historical percentage of Protestants. It reports the estimated long-run effect of the percentage of Protestants in 1517 ($\hat{\eta}_{1517}$), the estimated persistence of the percentage of Protestants from 1816 to 1871 ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of the percentage of Protestants. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

[a] Percentage blind in 1871, percentage deaf in 1871, and percentage insane in 1871.

[b] Percentage age below 10 years, percentage Jews, percentage females, percentage born in municipality, percentage of Prussian origin, average household size, log population size, population growth in 1867–1871 in percent, and percentage missing education information.

[c] Log distance to Berlin, latitude (in radians), longitude (in radians), latitude (in radians) × longitude (in radians), polish-speaking, percentage of labor force in mining, and percentage of population in urban areas.

[d] Combined district (entering as dummies for each category).

[e] Year when annexed by Prussia (entering as dummies for each category).

[f] Year when annexed by Prussia (entering as dummies for each category).

Table A.6: The Long-Run Effect of Protestantism on Distance to School

	Distance to School > 3 km							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Long-Run Effect ($\hat{\eta}_{1571}$)	-0.031***	-0.026*	-0.048**	-0.041	-0.039***	-0.054***	-0.027***	-0.031***
	(0.010)	(0.015)	(0.022)	(0.042)	(0.013)	(0.017)	(0.010)	(0.011)
Persistency of Endogenous Variable ($\hat{\delta}$)	1.005***	1.021***	1.032***	0.971***	1.013***	1.012***	1.000***	0.991***
	(0.011)	(0.016)	(0.024)	(0.051)	(0.014)	(0.013)	(0.011)	(0.007)
Conventional 2SLS Estimate	-0.031***	-0.023*	-0.040**	-0.048	-0.036***	-0.050***	-0.027***	-0.033***
	(0.010)	(0.013)	(0.016)	(0.046)	(0.012)	(0.014)	(0.009)	(0.011)
Control Variable Group 1[a]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variable Group 2[b]	No	Yes	Yes	Yes	No	No	No	No
Control Variable Group 3[c]	No	No	Yes	Yes	No	No	No	No
Control Variable Group 4[d]	No	No	No	Yes	No	No	No	No
Control Variable Group 5[e]	No	No	No	No	Yes	No	No	No
Control Variable Group 6[f]	No	No	No	No	No	Yes	No	No
Wald Test of $\hat{\delta} = 1$ p-value	0.692	0.195	0.185	0.572	0.329	0.349	0.966	0.220
First Stage F-Statistic (K-P) of Conventional	92.636	57.717	70.352	6.037	86.457	64.406	86.463	75.641
First Stage F-Statistic (K-P) of Persistence	84.758	51.550	58.377	5.840	78.009	58.308	80.037	71.258
Number of Observations	280	280	280	280	280	280	249	219

This table presents the results of a series of augmented IV-regressions of dummy variable indicating if the distance to school in 1871 is above 3 km on the historical percentage of Protestants. It reports the estimated long-run effect of the percentage of Protestants in 1517 ($\hat{\eta}_{1517}$), the estimated persistence of the percentage of Protestants from 1816 to 1871 ($\hat{\delta}$), and the conventional 2SLS-estimate of the effect of the percentage of Protestants. Furthermore, the table reports the results of a Wald test of the estimate of the persistence coefficient is equal to one as well as the first-stage F-statistics (Kleibergen-Paap) for conventional 2SLS regression and for the estimation of persistence. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors derived with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

[a]Percentage blind in 1871, percentage deaf in 1871, and percentage insane in 1871.

[b]Percentage age below 10 years, percentage Jews, percentage females, percentage born in municipality, percentage of Prussian origin, average household size, log population size, population growth in 1867–1871 in percent, and percentage missing education information.

[c]Log distance to Berlin, latitude (in radians), longitude (in radians), latitude (in radians) × longitude (in radians), polish-speaking, percentage of labor force in mining, and percentage of population in urban areas.

[d]Combined district (entering as dummies for each category).

[e]Year when annexed by Prussia (entering as dummies for each category).

[f]Year when annexed by Prussia (entering as dummies for each category).

Table A.7: Summary Statistics for the Institutions Application (N=56)

	Average	P25	P50	P75	S.D.
Log GDP per capita in 1990s	8.11	7.29	8.21	8.82	1.06
Constraint on Executive in 1990s	4.52	2.95	4.45	6.30	1.90
Constraint on Executive in 1960s	3.78	2.37	3.15	5.24	2.06
Constraint on Executive in 1900	2.30	1	1	3	2.17
Log Capped European Settler Mortality	4.47	4.24	4.36	5.50	0.94
Log Absolute Latitude	2.39	2.08	2.61	3.07	1.00

Table A.8: Summary Statistics for Protestantism Application (N=280)

	Mean (1)	P25 (2)	P50 (3)	P75 (4)	S.D. (5)
Average Household Size	4.79	4.52	4.79	5.05	0.36
City Population in 1500	0.00	0	0	0	0.00
Hanseatic City in 1517	0.06	0	0	0	0.23
Imperial City in 1517	0.06	0	0	0	0.23
Log Total Population Size	10.85	10.6	10.9	11.1	0.41
Latitude (in Radians)	90.44	88.9	90.0	91.7	2.16
Longitude (in Radians)	22.04	13.5	22.4	29.0	7.82
Longitude × Latitude (in Radians)	20.02	12.1	20.5	26.0	7.32
Monestaries in 1517 per Square Kilometer	0.04	0	0.0011	0.0038	0.44
Population Growth 1867–1871 (in Percentages)	1.40	-1.17	0.58	3.05	4.26
Percentage Age below 10	24.70	23.3	24.9	26.2	2.36
Percentage Blind in 1871	0.09	0.075	0.090	0.11	0.03
Percentage Born in Municipality	59.55	50.9	58.4	68.0	11.49
Percentage Deaf in 1871	0.09	0.065	0.083	0.12	0.04
Percentage Females	50.96	50.3	51.1	51.9	1.50
Percentage Insane in 1871	0.22	0.14	0.17	0.24	0.15
Percentage Jews	1.02	0.31	0.77	1.31	1.06
Percentage Literate in 1871	88.23	87.5	92.6	95.3	11.43
Percentage Missing Education Info	1.64	0.98	1.48	2.19	0.99
Percentage of Prussian Origin	99.35	99.2	99.7	99.9	1.01
Percentage Protestants in 1816	58.90	9.21	77.0	98.9	41.11
Percentage Protestants in 1871	58.56	12.6	72.5	97.7	39.75
School in 1517	0.05	0	0	0	0.22
University in 1517	0.03	0	0	0	0.17

C Additional Figures

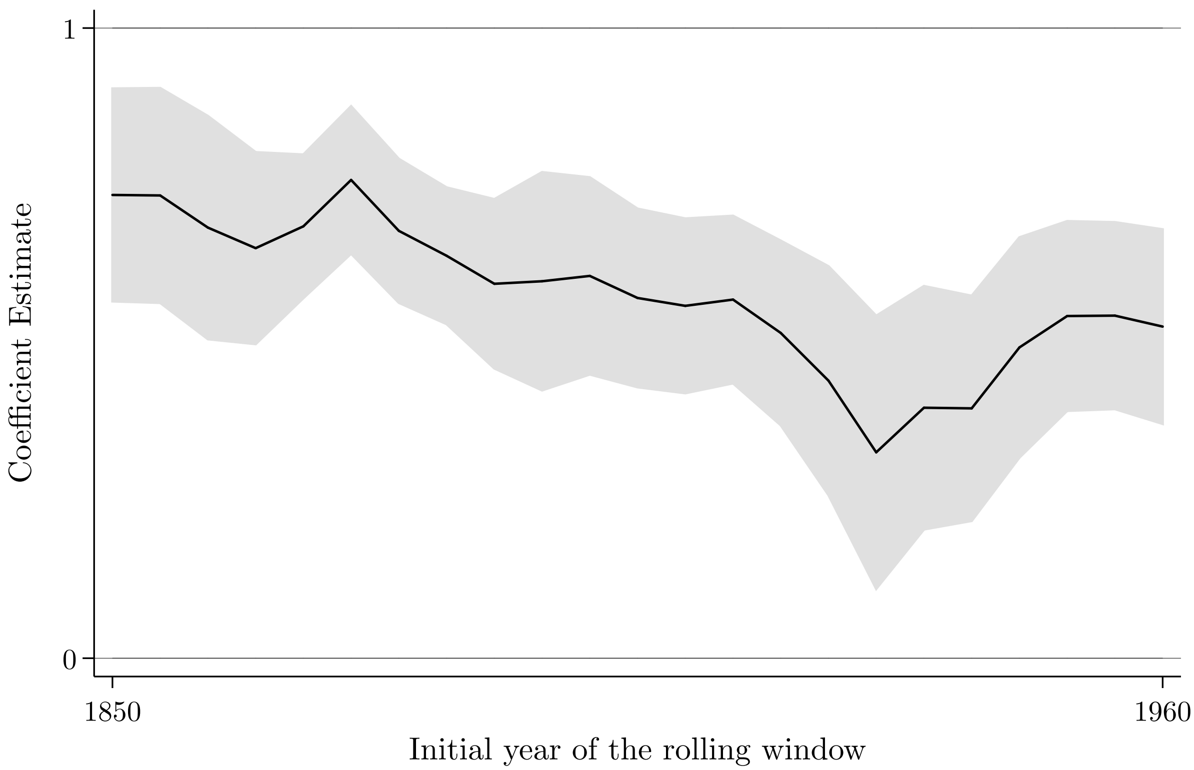

Figure A.1: This figure depicts the coefficient from five-year panel regressions of Constraints on the Executive on its lagged value in over period 1850–2013 with a 50-year regression window and a step size of five years, estimated with OLS. The sample is restricted to those 47 countries in the Polity IV database for which information on Constraints on the Executive exists for at least 75 percent of the years in the period 1850–2013. The regressions account for country and year fixed effects. Robust standard errors are used for the calculation of the confidence band.

Figure A.2: This figure depicts the coefficient from five-year panel regressions of Constraints on the Executive on its five-year lagged value in over period 1850–2013 with a 50-year regression window and a step size of five years, estimated with OLS. The sample is restricted to all 158 countries in the Polity IV database for which information on Constraints on the Executive exists for at least some of the years in the period 1850–2013. The regressions account for country and year fixed effects. Robust standard errors are used for the calculation of the confidence band.

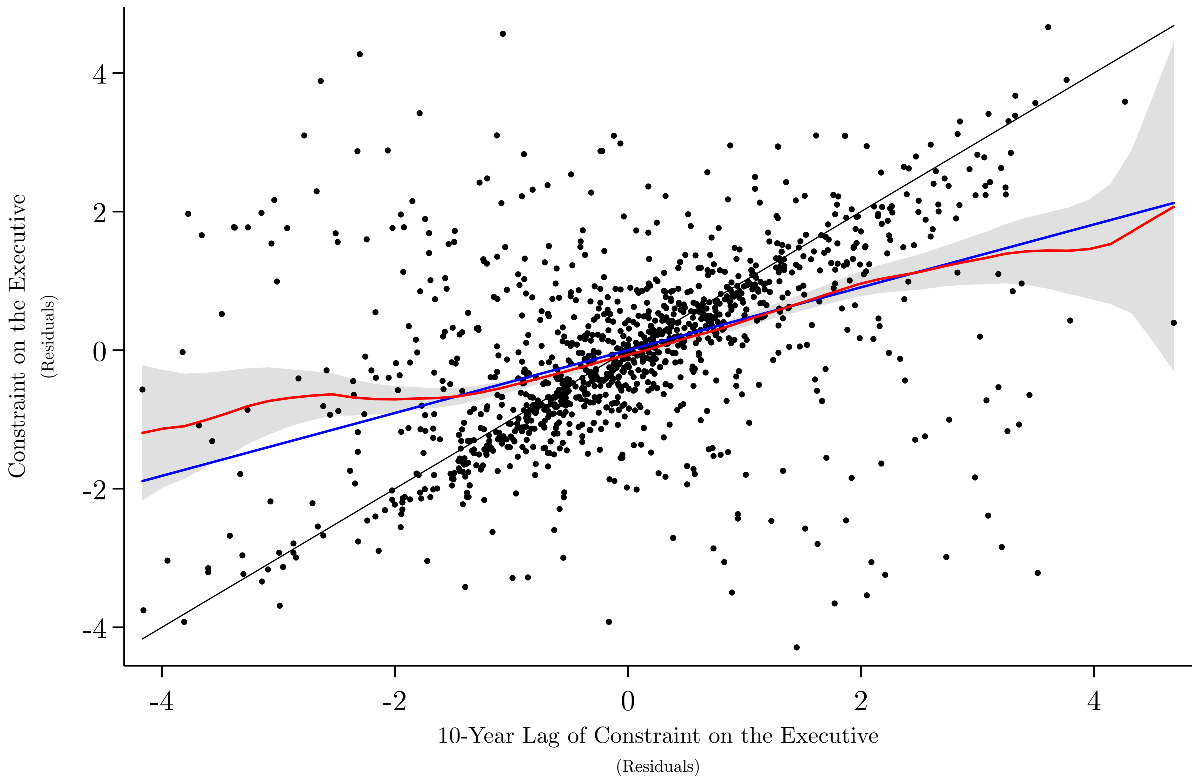

Figure A.3: Linear and flexible fits of δ using a 10-year lagged panel data model accounting for country and year fixed effects. The black line represents $\delta = 1$. The blue line represents the fit from a linear regression model using a linear specification. The red line represents a flexible fit from a kernel-weighted local-mean smoothing. The shaded area represents the 95 percent confidence bounds of the flexible fit.

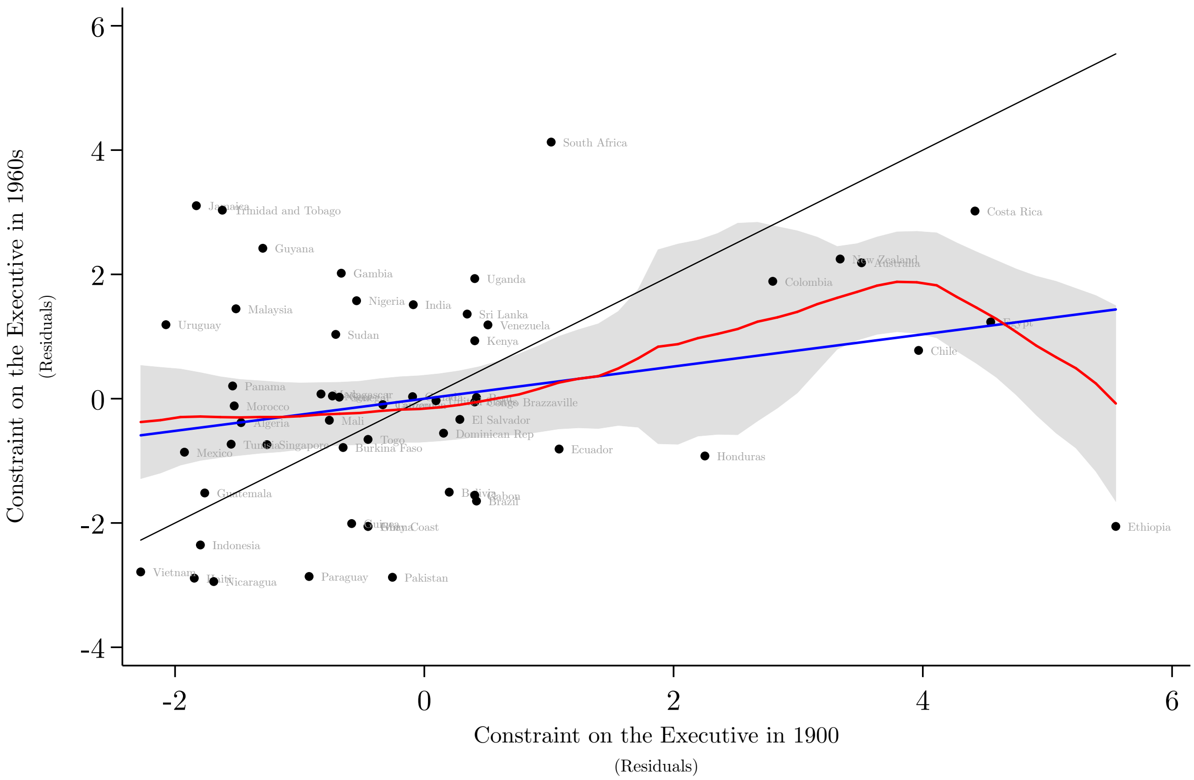

Figure A.4: Linear and flexible fits of δ for the cross sectional data underlying the institutions analysis. The black line represents $\delta = 1$. The blue line represents the fit from a linear regression model using a linear specification. The red line represents a flexible fit from a kernel-weighted local-mean smoothing. The shaded area represents the 95 percent confidence bounds of the flexible fit.

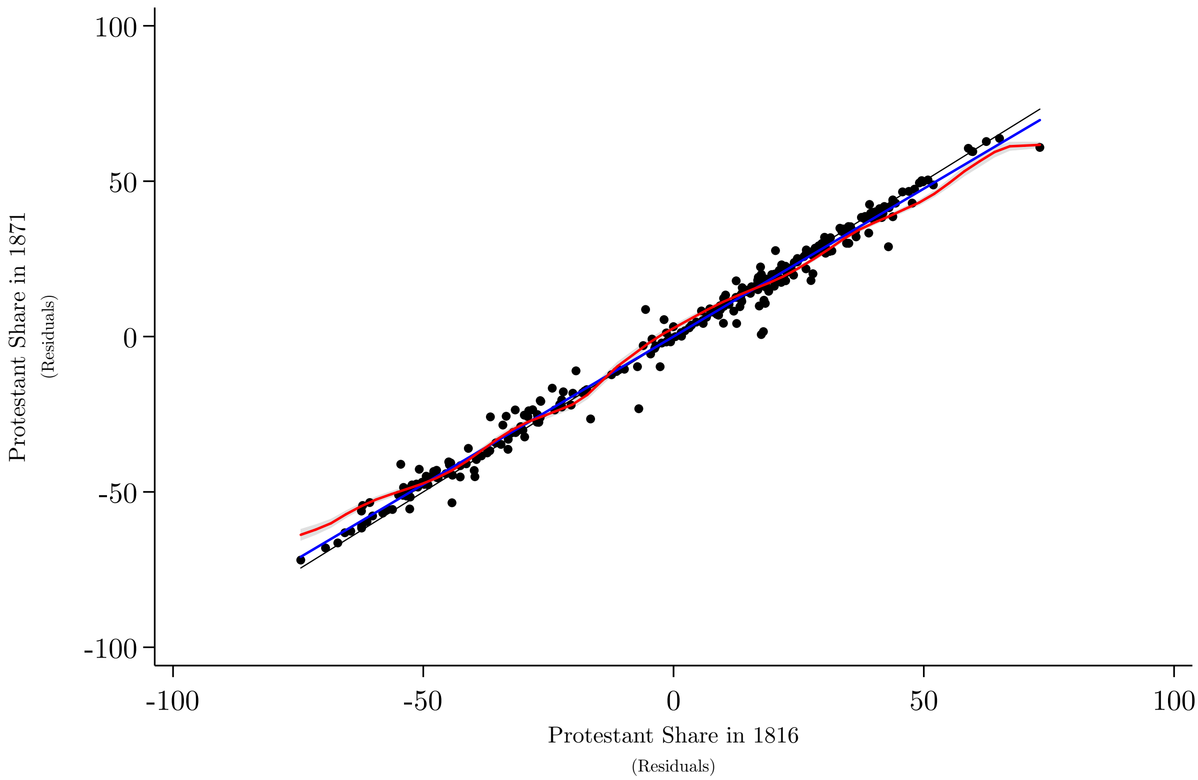

Figure A.5: Linear and flexible fits of δ for the cross sectional data underlying the Protestantism analysis. The blue line represents the fit from a linear regression model using a linear specification. The red line represents a flexible fit from a kernel-weighted local-mean smoothing. The shaded area represents the 95 percent confidence bounds of the flexible fit.

References

Acemoglu, D. (2010): "Theory, General Equilibrium, and Political Economy in Development Economics," *Journal of Economic Perspectives*, 17–32.

Acemoglu, D., F. A. Gallego, and J. A. Robinson (2014): "Institutions, Human Capital, and Development," *Annual Review of Economics*, 6, 875–912.

Acemoglu, D. and S. Johnson (2005): "Unbundling institutions," *Journal of Political Economy*, 113, 949–995.

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

Acemoglu, D., S. Johnson, and J. A. Robinson (2005): "Institutions as a fundamental cause of long-run growth," *Handbook of Economic Growth*, 1, 385–472.

Acemoglu, D., S. Johnson, and J. A. Robinson (2012): "The colonial origins of comparative development: An empirical investigation: Reply," *American Economic Review*, 102, 3077–3110.

Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008): "Income and democracy," *American Economic Review*, 98, 808–842.

Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2009): "Reevaluating the modernization hypothesis," *Journal of Monetary Economics*, 56, 1043–1058.

Albouy, D. Y. (2012): "The colonial origins of comparative development: An empirical investigation: comment," *American Economic Review*, 102, 3059–3076.

Alesina, A. and P. Giuliano (2013): "Family ties," *Handbook of Economic Growth*, 2, 177.

Alesina, A., P. Giuliano, and N. Nunn (2011): "Fertility and the plough," *American Economic Review*, 101, 499–503.

Alesina, A., J. Harnoss, and H. Rapoport (2016): "Birthplace diversity and economic prosperity," *Journal of Economic Growth*, 21, 101–138.

Andersen, T. B., C.-J. Dalgaard, and P. Selaya (2016): "Climate and the emergence of global income differences," *Review of Economic Studies*.

Ang, J. B. (2013): "Institutions and the long-run impact of early development," *Journal of Development Economics*, 105, 1–18.

Ashraf, Q. and O. Galor (2013): "The "out of Africa" hypothesis, human genetic diversity, and comparative economic development," *American Economic Review*, 103, 1–46.

Ashraf, Q. and O. Galor (2016): "The macrogenoeconomics of comparative development," *Working Paper*.

Auer, R. A. (2013): "Geography, institutions, and the making of comparative development," *Journal of Economic Growth*, 18, 179–215.

Banerjee, A. and L. Iyer (2005): "History, institutions, and economic performance: The legacy of colonial land tenure systems in India," *American Economic Review*, 95, 1190–1213.

Becker, S. O., F. Cinnirella, E. Hornung, and L. Woessmann (2014): "iPEHD – the ifo Prussian economic history database," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 47, 57–66.

Becker, S. O., F. Cinnirella, and L. Woessmann (2010): "The trade-off between fertility and education: evidence from before the demographic transition," *Journal of Economic Growth*, 15, 177–204.

Becker, S. O. and L. Woessmann (2009): "Was Weber wrong? A human capital theory of Protestant economic history," *Quarterly Journal of Economics*, 124, 531–596.

Benhabib, J., A. Corvalan, and M. M. Spiegel (2011): "Reestablishing the income-democracy nexus," *NBER Working Paper*.

Bruhn, M. and F. A. Gallego (2012): "Good, bad, and ugly colonial activities: Do they matter for economic development?" *Review of Economics and Statistics*, 94, 433–461.

Cervellati, M., F. Jung, U. Sunde, and T. Vischer (2014): "Income and democracy: Comment," *American Economic Review*, 104, 707–719.

Cervellati, M. and U. Sunde (2015): "The economic and demographic transition, mortality, and comparative development," *American Economic Journal: Macroeconomics*, 7, 189–225.

Comin, D., W. Easterly, and E. Gong (2010): "Was the Wealth of Nations determined in 1000 BC?" *American Economic Journal: Macroeconomics*, 2, 65–97.

Deaton, A. (2010a): "Instruments, randomization, and learning about development," *Journal of Economic Literature*, 424–455.

Deaton, A. (2010b): "Understanding the mechanisms of economic development," *Journal of Economic Perspectives*, 3–16.

Dell, M. (2010): "The persistent effects of Peru's mining mita," *Econometrica*, 78, 1863–1903.

Do, Q.-T., A. A. Levchenko, and C. Raddatz (2016): "Comparative advantage, international trade, and fertility," *Journal of Development Economics*, 119, 48–66.

Dollar, D. and A. Kraay (2003): "Institutions, trade, and growth," *Journal of Monetary Economics*, 50, 133–162.

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

Easterly, W. and R. Levine (2003): "Tropics, germs, and crops: how endowments influence economic development," *Journal of Monetary Economics*, 50, 3–39.

Feyrer, J. (2009a): "Distance, trade, and income-the 1967 to 1975 closing of the suez canal as a natural experiment," *NBER Working Paper*.

Feyrer, J. (2009b): "Trade and Income–Exploiting Time Series in Geography," *NBER Working Paper*.

Feyrer, J. and B. Sacerdote (2009): "Colonialism and modern income: Islands as natural experiments," *Review of Economics and Statistics*, 91, 245–262.

Frankel, J. A. and D. Romer (1999): "Does trade cause growth?" *American Economic Review*, 379–399.

Galor, O. and Ö. Özak (2016): "The Agricultural Origins of Time Preference," *American Economic Review*.

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

Gorodnichenko, Y. and G. Roland (2011): "Which dimensions of culture matter for long-run growth?" *American Economic Review*, 101, 492–498.

Gorodnichenko, Y. and G. Roland (2017): "Culture, institutions and the wealth of nations," *Review of Economics and Statistics*, 99, 402–416.

Guiso, L., P. Sapienza, and L. Zingales (2006): "Does culture affect economic outcomes?" *Journal of Economic Perspectives*, 20, 23–48.

Heid, B., J. Langer, and M. Larch (2012): "Income and democracy: Evidence from system GMM estimates," *Economics Letters*, 116, 166–169.

Iyer, L. (2010): "Direct versus indirect colonial rule in India: Long-term consequences," *Review of Economics and Statistics*, 92, 693–713.

Michalopoulos, S. and E. Papaioannou (2013): "Pre-colonial ethnic institutions and contemporary African development," *Econometrica*, 81, 113–152.

Naritomi, J., R. R. Soares, and J. J. Assunção (2012): "Institutional development and colonial heritage within Brazil," *Journal of Economic History*, 72, 393–422.

Nunn, N. (2008): "The long-term effects of Africa's slave trades," *Quarterly Journal of Economics*, 123, 139–176.

Nunn, N. (2014): "Historical development," *Handbook of Economic Growth*, 2, 347–402.

Putterman, L. and D. Weil (2010): "Post-1500 population flows and the long run determinants of economic growth and inequality," *Quarterly Journal of Economics*, 125, 1627.

Rodrik, D., A. Subramanian, and F. Trebbi (2004): "Institutions rule: The primacy of institutions over geography and integration in economic development," *Journal of Economic Growth*, 9, 131–165.

Spolaore, E. and R. Wacziarg (2013): "How deep are the roots of economic development?" *Journal of Economic Literature*, 51, 325–69.

Tabellini, G. (2010): "Culture and institutions: Economic development in the regions of Europe," *Journal of the European Economic Association*, 8, 677–716.

Weber, M. (1958): *The Protestant Ethic and the Spirit of Capitalism*.