Journal Pre-proof

Historical instruments and contemporary endogenous regressors

Gregor Casey, Marc Klemp

Please cite this article as: Casey, G., Klemp, M., Historical instruments and contemporary endogenous regressors, *Journal of Development Economics* (2021), doi: https://doi.org/10.1016/j.jdeveco.2020.102586.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Historical Instruments and Contemporary Endogenous Regressors[*]

Gregory Casey   Marc Klemp

November 2020

Abstract

We provide a simple framework for interpreting instrumental variable regressions when there is a gap in time between the impact of the instrument and the measurement of the endogenous variable, highlighting a particular violation of the exclusion restriction that can arise in this setting. In the presence of this violation, conventional IV regressions do not consistently estimate a structural parameter of interest. Building on our framework, we develop a simple empirical method to estimate the long-run effect of the endogenous variable. We use our bias correction method to examine the role of institutions in economic development, following Acemoglu et al. (2001). We find long-run coefficients that are smaller than the coefficients from the existing literature, demonstrating the quantitative importance of our framework.

Keywords Long-Run Economic Development, Instrumental Variable Regression

JEL Classification Codes C10, C30, O10, O40

1   Introduction

In empirical economic research, it is often difficult to assign causality, especially when investigating processes that unfold over time. At the same time, a growing literature convincingly argues that historical events affect contemporary economic outcomes.[1] To find sources of (quasi-)random variation, therefore, researchers sometimes turn to historical events. In the context of instrumental variables (IV) regressions, this can result in cases where there is a significant gap in time between the initial impact of the instrument and the measurement of the endogenous variable (e.g., Levine et al., 2000; Acemoglu et al., 2001; Tabellini, 2010).

We study the interpretation of these IV regressions. To start, we provide a simple framework for analyzing these situations, highlighting a particular violation of the exclusion restriction that can only arise when a time gap exists. This violation occurs when past, unmeasured values of the endogenous variable exert an influence on the outcome variable that is not mediated by the contemporary, measured value of the endogenous variable. This violation occurs even when the historical variable would be a valid instrument for the historical value of the endogenous variable.

Our framework demonstrates that conventional IV regressions with a time gap estimate the ratio of the long-run effect and the persistence of the endogenous variable. The long-run effect is the parameter that would be estimated if the endogenous variable was measured at the same time as the initial impact of the instrument. We use 'persistence' to denote the causal effect of the historical level (time period of the impact of instrument) of the endogenous variable on the contemporary level (time period of the outcome variable) of the endogenous variable.

Based on these results, we extend our simple framework to demonstrate how to estimate long-run causal effects under common data availability constraints. Our empirical approach requires jointly estimating two equations using a single instrument. One equation estimates the conventional IV regression with a time gap. The other equation estimates the persistence of the endogenous variable between two intermediate points in time, which can be combined with structural assumptions to infer persistence over the whole period of interest. The estimates from the latter regression are then used to correct the bias in the former regression.[2]

We use our new bias correction method to re-examine the relationship between institutions and economic development, 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 increases 1990s income per capita by approximately 0.7 standard deviations. While sizable, this effect size is approximately 16% of the coefficient generated by the conventional IV regression, indicating that our proposed method is quantitatively important for estimating long-run effects. We use panel data on institutions to validate the key assumptions of our bias correction

---

[1]For overviews focusing on economic development, see Spolaore and Wacziarg (2013), Nunn (2014), and Michalopoulos and Papaioannou (2020).

[2]We also show how to combine our framework with the results of Conley et al. (2012) to better understand the contemporaneous relationship between the outcome variable and the endogenous regressor of interest.

method. In addition, we discuss other papers that have a time gap in time between the initial impact of the instrument and the measurement of the endogenous regressor.[3]

Finally, we discuss the implications of our framework for future applied work. We start by describing the practical steps to implementing our method. When it is not possible to implement our bias correction method due to lack of data or failure of the underlying assumptions, focusing on the reduced form and first stage regressions separately can still generate important insights, even though the IV regression does not consistently estimate a structural parameter of interest. When trying to estimate long-run effects, the issue we highlight is fundamentally about mis-measurement of the endogenous variable. As a result, the collection of historical data can also overcome these issues.

The remainder of the paper proceeds as follows. In section 2, we present our framework and main analytic results. In section 3, we present the empirical application. In section 4, we discuss additional examples, and in section 5, we discuss practical considerations for applied research. Section 6 concludes.

2   Framework and Analytic Results

2.1   Interpreting IV regressions with historical instruments and contemporary regressors

Figure 1 provides a representation of our framework.[4] We start by just considering the top row (i.e., we ignore A_C). Our endogenous explanatory variable of interest is X, and Y_C 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 period in which Y_C is measured. We assume that Z_H would be a valid instrument for X_H, but that X_H is unobserved. This is a common data availability constraint. A data generating process of this form is often implicitly assumed to underlie regressions using historical instruments and contemporary endogenous regressors.

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_C in the contemporary period, then X_H should in general also affect Y_H (not shown) 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_H are persistent — then there will be a causal effect of X_H on Y_C that is not mediated by X_C. We represent this link using the variable A_C, which is a reduced form representation of a more complicated dynamic process. In most applications, it is unlikely that all

---

[3]We also discuss the related case where historical values of the endogenous variables are used as the instrument. These regressions have identical issues of interpretation as those falling within our main framework, but long-run effects can be estimated without our bias correction method.

[4]When abstracting from the time dimension, the data generating process considered here is similar to that of Dippel et al. (2017), who assume that the endogenous regressor of interest is observable and decompose its causal effect between direct and mediated channels. Our goal, by contrast, is to analyze the case where X_H is unobserved, as is frequently the case when using historical instruments and contemporary endogenous regressors.
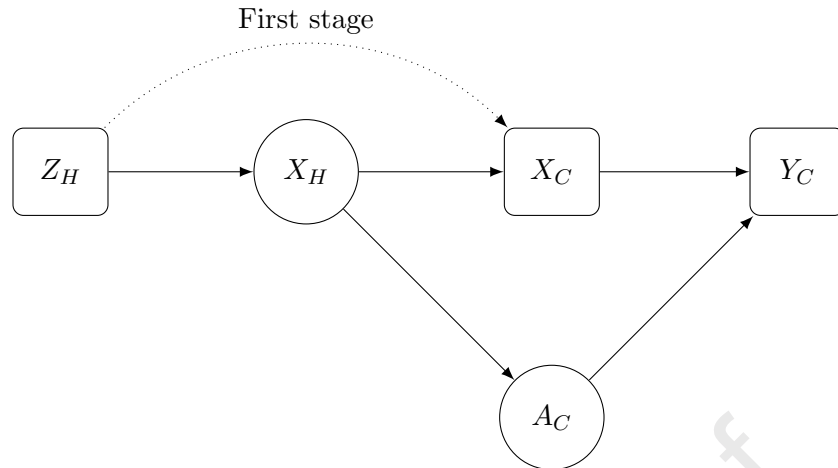
First stage

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.

components of A_C are observed.[5] Thus, we assume that A_C is unobserved. We will refer to A_C as an 'alternative channel.'[6]

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_H is settler mortality, Y_C 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_C variable, which is consistent with the empirical findings and interpretation presented in their paper.[7] The A_C variable could be physical or human capital, technology, or culture.

Equations (1)–(4) represent the data generating process algebraically:[8]

$$X_{H,i} = \psi Z_{H,i} + \varepsilon_{X_H,i}, \tag{1}$$

$$X_{C,i} = \delta X_{H,i} + \varepsilon_{X_C,i}, \tag{2}$$

$$A_{C,i} = \gamma X_{H,i} + \varepsilon_{A,i}, \tag{3}$$

$$Y_{C,i} = \beta_1 X_{C,i} + \beta_2 A_{C,i} + \varepsilon_{Y,i}. \tag{4}$$

In standard microeconomic settings, instrumental variables are used to estimate the contemporaneous causal effect of X, $\frac{\partial Y_C}{\partial X_C} = \beta_1$. We are interested in research designs that use $Z_{H,i}$ as an

---

[5]Observing part, but not all, of the A variable would further complicate the interpretation of the regression.

[6]For the remainder of the paper, when we refer to an A_C variable or an alternative channel, we focus on the case where $\gamma \neq 0$ and $\beta_2 \neq 0$. Appendix section A.1 analyzes the case without A_C.

[7]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).

[8]To economize on notation, we do not assume that the ε terms are mean zero, implying that they also capture the constant term for each equation.

instrument for $X_{C,i}$ and estimate equations of the form

$$Y_{C,i} = b_0 + b_1 X_{C,i} + \tilde{\varepsilon}_i. \tag{5}$$

Given our structural framework, $\tilde{\varepsilon}_i = \beta_2 A_{C,i} + \varepsilon_{Y,i}$. Instruments are used because of concerns that $Cov(X_C, \tilde{\varepsilon}) \neq 0$, due to reverse causality or omitted variables. Here, we separately specify A_C, which is a particular type of (structural) omitted variable. When explicitly writing out the underlying model, it is clear that this regression will not consistently estimate β_1, because $Cov(Z_H, A_C) \neq 0 \Rightarrow Cov(Z_H, \tilde{\varepsilon}_Y) \neq 0$.

While this is obviously an econometric problem, it is not clear that β_1 is always the true parameter of interest. Instead, researchers often loosely interpret (5) as providing information about the long-run impact of historical factors on contemporary outcomes. As a result, our reading is that the long-run causal effect of X, $\eta \equiv \frac{\partial Y_C}{\partial X_H}$, is often a key parameter. A little algebra gives

$$Y_{C,i} = (\delta\beta_1 + \beta_2\gamma) X_{H,i} + \mu_i, \tag{6}$$

where $\mu_i = \varepsilon_{Y,i} + \beta_1\varepsilon_{X_C,i} + \beta_2\varepsilon_{A,i}$. So, $\eta = \delta\beta_1 + \beta_2\gamma$. 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.

When discussing the validity of the instrument, Z_H, the literature focuses on the fact that it exogenously shifts X_H. We assume, therefore, that[9]

$$Cov(Z_H, \mu) = Cov(Z_H, \varepsilon_Y) = Cov(Z_H, \varepsilon_{X_C}) = Cov(Z_H, \varepsilon_A) = 0. \qquad \text{(Assumption 1)}$$

With these assumptions, estimation of (5) with Z_H as an instrument yields:[10]

$$\begin{aligned}
\text{plim } \hat{b}_1^{IV} &= \frac{Cov(Y_C, Z_H)}{Cov(X_C, Z_H)} \tag{7} \\
&= \frac{\beta_1 Cov(X_C, Z_H) + \beta_2 Cov(A_C, Z_H)}{Cov(X_C, Z_H)} \tag{8} \\
&= \beta_1 + \frac{\beta_2\gamma}{\delta} = \frac{\eta}{\delta}. \tag{9}
\end{aligned}$$

We can see that 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 inability of the regression to consistently estimate η is a measurement problem. The endogenous variable is X_C

---

[9]In the context of Acemoglu et al. (2001), it is important to note that Assumption 1 rules out the existing critique raised by Glaeser et al. (2004), who argue that the initial impact of settler mortality works through other channels, like human capital (see also, Easterly and Levine (2016)). Thus, they assume that $Cov(Z_H, \mu) \neq 0$. In contrast, our framework accepts the general premise of Acemoglu et al. (2001), but investigates the role of the underlying dynamics in understanding the cross-sectional regression coefficients. See Auer (2013) for an approach to addressing these older critiques.

[10]To simplify the algebra, note that $Cov(A_C, Z_H) = \gamma Cov(X_H, Z_H)$ and $Cov(X_C, Z_H) = \delta Cov(X_H, Z_H)$.

instead of X_H, and the degree of (non-classical) measurement error is given by persistence (δ). A large conventional regression coefficient may indicate either a large impact of X_H or low persistence in X.

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 $\delta = 1$. 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 magnitude of the long-run effect of X on Y. As demonstrated in appendix section A.1.1, the relationship between the regression coefficient and η is unchanged if the A_C variable is excluded from the system.

These results establish 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, however, X_H is not observed. Thus, we need to combine the cross-sectional regression with an estimate of δ. In section 2.2, we demonstrate how to estimate η in this manner.

Estimating contemporaneous relationships. — In studies using historical instruments and contemporary endogenous regressors, we believe that η is often the fundamental parameter of interest. Depending on the question, however, researchers may be more interested in contemporaneous relationships. Before turning to our new method, therefore, it is helpful to consider the implications of our framework for estimating β_1.

According to equation (9), $\text{plim}\,\hat{b}_1^{IV} = \beta_1 + \frac{\beta_2\gamma}{\delta}$. Unsurprisingly, the inconsistency is affected by the strength and sign of the A_C channel. Specifically, γ gives the effect of X_H on A_C, and β_2 gives the effect of A_C on Y_C. If either of these effects has a large magnitude, the IV regression is likely to give a misleading estimate of β_1. As in the case of the long-run effect, the inconsistency in estimating β_1 also depends on the degree on persistence.[11] In particular, the absolute value of the inconsistency is small when the persistence is large. Intuitively, if δ is large, then the variation in X_H generated by Z is small compared to the variation in X_C generated by Z. As a result, the impact of X_H on Y_C through alternate channels (i.e., the violation of the exclusion restriction) will not have a large effect on the estimated coefficient.

To formally bound estimates of β_1, researchers can use the 'plausibly exogenous' instruments framework of Conley et al. (2012). From equations (1)–(4), it is straightforward to derive

$$Y_{C,i} = \beta_1 X_{C,i} + \beta_2\gamma\psi Z_i + \tilde{\mu}_i, \tag{10}$$

and

$$X_{C,i} = \delta\psi Z_{H,i} + (\delta\varepsilon_{X_H,i} + \varepsilon_{X_C,i}), \tag{11}$$

where $\tilde{\mu}_i = (\beta_2\gamma\varepsilon_{X_H,i} + \beta_2\varepsilon_{A,i} + \varepsilon_{Y,i})$. Conley et al. (2012) show that conventional IV regressions can bound β_1 when combined with assumptions about the prior distribution or support of $\beta_2\gamma\psi$.

---

[11]In principle, the inconsistency could also depend on the sign of δ, but we believe that negative persistence is unlikely to be important in many empirical settings.

Our framework attaches a structural interpretation to $\beta_2\gamma\psi$, which makes it easier to ground these assumptions in economic theory or available empirical evidence. In our framework, both the causal effect of the endogenous variable (β_1) and the violation of the exclusion restriction both stem from the initial impact of Z on X_H (ψ). In other words, holding all else equal, the violation of the exclusion restriction is more problematic when the first stage relationship is strong.

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, we extend our framework to allow for more than two periods:

$$X_{t,i} = \kappa_{X_t} + \delta X_{t-1,i} + \varepsilon_{X_t,i}, \ \forall \ t = 1\ldots C, \ t \neq H, \tag{12}$$

$$X_{H,i} = \kappa_{X_H} + \delta X_{H-1,i} + \psi Z_{H,i} + \varepsilon_{X_H,i}, \tag{13}$$

$$A_{C,i} = \kappa_A + \gamma X_{H,i} + \varepsilon_{A,i}, \tag{14}$$

$$Y_{C,i} = \kappa_Y + \beta_1 X_{C,i} + \beta_2 A_{C,i} + \varepsilon_{Y,i}. \tag{15}$$

Now, X initially follows a simple law of motion given by (12). In some period H, X_H is shocked by Z_H. 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] We assume that

$$Cov(Z_H, \varepsilon_{Y_C}) = Cov(Z_H, \varepsilon_{A_C}) = Cov(Z_H, \varepsilon_{X_t}) = 0 \ \forall t. \tag{Assumption 2}$$

Our method requires that X be observed at two different points is time. We label these time periods T and $T - Q$, where $0 < Q < T$. By assumption, we do not observe X_H, implying that $T - Q > H$. Now, we solve 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 (12) recursively:

$$X_{T,i} = \kappa_{X_T} + \delta X_{T-1,i} + \varepsilon_{X_T,i} = \tilde{\kappa}_X + \delta^Q X_{T-Q,i} + \tilde{\varepsilon}_{X,i}, \tag{16}$$

where $\tilde{\kappa}_X = \sum_{k=0}^{Q-1} \delta^k \kappa_{X_{T-k}}$ is a constant, and $\tilde{\varepsilon}_{X,i} = \sum_{k=0}^{Q-1} \delta^k \varepsilon_{X_{T-k},i}$ is an observation-specific error term.

Now, consider the IV regression equation

$$X_{T,i} = a_0 + a_1 X_{T-Q,i} + a_{2,i}, \tag{17}$$

with Z_H is an instrument for X_{T-Q}. There is no violation of the exclusion restriction in this case, and according to (16), the estimation yields

$$\text{plim} \ \hat{a}_1 = \delta^Q. \tag{18}$$

---

[12]This result can be generalized to a time-varying δ, provided that the nature of the time dependence is known.

This is the aggregate degree of persistence over Q periods.

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

$$Y_{C,i} = \tilde{\beta}_0 + (\beta_1 \delta^{C-H} + \beta_2 \gamma) X_{H,i} + \tilde{\epsilon}_i, \tag{19}$$

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

$$Y_{C,i} = b_0 + b_1 X_{C,i} + b_{2,i}, \tag{20}$$

where Z_H is an instrument for X_C. Similar to our results from section 2.1, this regression yields[13]

$$\operatorname{plim} \hat{b}_1 = \frac{\beta_1 \delta^{C-H} + \beta_2 \gamma}{\delta^{C-H}} = \frac{\eta}{\delta^{C-H}}. \tag{21}$$

Here, δ^{C-H} is the total degree of persistence from the time of the shock to the time that Y is measured.

To solve for η, we simply combine the results from estimating equations (17) and (20),

$$\operatorname{plim} \hat{a}_1 = \delta^Q$$

and

$$\operatorname{plim} \hat{b}_1 = \frac{\eta}{\delta^{C-H}}.$$

Putting these together yields

$$\eta = (\operatorname{plim} \hat{b}_1)(\operatorname{plim} \hat{a}_1)^{\frac{C-H}{Q}}. \tag{22}$$

To estimate η, we first estimate equations (17) and (20) via instrumental variables in order to obtain $\hat{b}_1$ and $\hat{a}_1$.[14] Then, we combine the two regression coefficients using the nonlinear function in (22). To construct confidence intervals, we apply the delta method.

It is worth noting two strong 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 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 for at least three points in time. In the empirical application, we investigate the validity of these assumptions using panel data.

---

[13]To see this, plug X_H and X_C into equation (16). Combined with equations (13), (14), and (15), this is just the original system of equations from section 2.1, except that δ is persistence over one period and total persistence is given by δ^{C-H}.

[14]These equations can be jointly estimated, e.g., via stacked 2SLS regressions or multiple-equation instrumental variable GMM.

3 Empirical Application: Institutions and Income per Capita

In our empirical application, we examine the effect of institutions on economic development, following Acemoglu et al. (2001). We choose this application for several reasons. First, this is likely the most prominent paper using historical instruments for contemporary endogenous regressors, and many important papers in the comparative development literature follow the methodology developed in the article. 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 7). 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 using our method to estimate η and in validating the assumptions.

3.1 Main Results

Our measure of institutions, 'Constraints on the Executive,' comes from the Polity5 dataset. It measures the limits to executive power on a seven point scale that increases in the level of constraints. This is the preferred measure of institutions in the literature (Glaeser et al., 2004; Acemoglu et al., 2005). The outcome variable is the natural log of income per capita in 1990, and the instrument is settler mortality.[15] Since settler mortality may be correlated with region-specific factors, such as 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] Appendix Table A.1 provides summary statistics.

We apply our bias correction method from section 2.2 to estimate the long-run effect of institutions on economic development. To do so, we simultaneously estimate two sets of equations via stacked 2SLS. In the first set, we estimate the cross-sectional relationship between contemporary institutions and contemporary income per capita via equation (17):

$$\text{GDPpc}_{C,i} = b_0 + b_1\text{Inst}_{C,i} + \text{error}, \tag{23}$$
$$\text{Inst}_{C,i} = d_0 + d_1\text{SettMort}_{H,i} + \text{error}, \tag{24}$$

where $GDPpc$ is income per capita in country i, $Inst$ is a measure of the quality of institutions, $SettMort$ is settler mortality, C represents the contemporary period, and H represents the historical period. Following the original research, we will take the contemporary period to be 1995.[17] The timing of the initial shock is difficult to determine exactly and likely differs across countries. We

---

[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 regressions. The uncapped settler mortality variable is obtained directly from Acemoglu et al. (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]In practice, we average outcome data over the period 1990-1999.

take a conservative approach and use $H = 1800$. 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. In the second set of equations, we estimate the persistence of institutions via equation (20):

$$\text{Inst}_{T,i} = a_0 + a_1 \text{Inst}_{T-Q,i} + \text{error}, \tag{25}$$

$$\text{Inst}_{T-Q,i} = f_0 + f_1 \text{SettMort}_{H,i} + \text{error}, \tag{26}$$

Here, T and $T - Q$ are two points in time which institutions are measured, where $T \in (H, C]$ and $Q \in (0, C - H)$. We estimate equation (20) using Constraints on the Executive data for the period 1900–1990s ($T = 1995$ and $T - Q = 1900$).[18]. Then, to estimate $\eta_{1800} \equiv \frac{\partial \text{GDPpc}_{1995}}{\partial \text{Inst}_{1800}}$, we use equation (22),

$$\hat{\eta}_{1800} = \hat{b}_1 \cdot \hat{a}_1^{\frac{1995-1800}{1995-1900}}, \tag{27}$$

and use the delta method to compute standard errors.[19]

Table 1 presents the results from our analysis.[20] Column 1 examines the case without any control variables. We estimate that raising Constraints on the Executive in 1800 by one point on the 7-point index increases contemporary income per capita by 0.38 log points. This implies that increasing constraints from the lowest possible score (1) to the highest possible score (7) increases log 1990 income per capita by approximately 2.3 standard deviations. While this is an economically significant effect, the estimated long-run coefficient is only 55% as large as the conventional IV estimate. Thus, accounting for the persistence in the endogenous explanatory variable is quantitatively important for the estimation of the long-run effect.

We find that the conventional IV regression overestimates the long-run effect. This occurs because institutions are less than perfectly persistent ($\delta < 1$). An increase in Constraints on the Executive in the contemporary period of one unit, therefore, corresponds to an increase in the 1800 measure of institutions of more than one unit. In the next subsection, we use panel data to corroborate our finding that δ is significantly smaller than one.

Settler mortality may be correlated with other geographic factors that affect contemporary income per capita, creating a classic violation of the exclusion restriction. Thus, the remaining columns of the table adds latitude and World Bank region fixed effects. 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 log 1990 income per capita by approximately 0.7 standard deviations. This long-run coefficient is about 16% as large as the conventional IV estimate.

---

[18]Our analysis of the data revealed a decline in persistence in the post-1960 period. We therefore also estimate equation (20) using data for the period 1900–1960s (Table A.2)

[19]We use the same sample when estimating both sets of regressions. In the presence of heterogenous treatment effects, the results should be interpreted as the local average treatment effect for the complier group affected by settler mortality.

[20]Tables A.4 and A.3 show that the results of Table 1 and Table A.2 are robust to the use of GDP per capita in 2019 as an alternative measure of contemporary income.

9

Table 1: The Long-Run Effect of Institutions on Income Per Capita

	Log GDP per capita in 1990s		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.381**	0.114	0.118
	(0.163)	(0.144)	(0.155)
Persistency of Endogenous Variable ($\hat{a}_1$)	0.751***	0.448*	0.413
	(0.164)	(0.263)	(0.258)
Conventional 2SLS Estimate ($\hat{b}_1$)	0.687***	0.593***	0.725***
	(0.153)	(0.213)	(0.216)
World Region Fixed Effects	No	Yes	Yes
Absolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.128	0.036	0.023
First Stage F-Statistic (K-P) of Conventional	27.050	4.241	3.881
First Stage F-Statistic (K-P) of Persistence	26.367	12.587	11.135
Number of Observations	56	56	56

This table reports the estimated long-run effect of constraints on the executive in 1800 on log GDP per capita in the 1990s ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1990s ($\hat{a}_1$), and the conventional 2SLS-estimate of the effect of constraints on the executive in 1990s on log GDP per capita in the 1990s ($\hat{\beta}_1$). It also reports the results of a Wald test for the null hypothesis that the persistence coefficient is equal to one. In addition, it reports the first-stage F-statistics (Kleibergen-Paap) for the conventional 2SLS regression and the persistence regression. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors calculated with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

3.2 Assessment of Imperfect, Constant, and Linear Persistence

In the section, we use panel data to support our findings and examine the key assumptions that we impose in order to estimate η. We employ the panel-model analog of equation (17):

$$\text{Inst}_{i,t} = \alpha_i + \nu_t + \delta \text{Inst}_{i,t-1} + \varepsilon_{i,t}, \tag{28}$$

where Inst is a measure of institutions, ν_t is a time period fixed effect, and α_i is a country fixed-effect. We use the Constraints on the Executive data from the Polity5 version 2018 dataset that covers the period 1800–2018 (Marshall and Gurr, 2020). Comparing equation (28) with equation (16) in section 2.1 shows that δ in equation (28) is the relevant measure of institutional persistence. We run the regressions with 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.[21] Unlike the main analysis, we do not have an 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

---

[21]Since there is missing data for some period-country pairs, averaging increases the sample size. It may also help counter attenuation bias.

the same direction, biasing our estimate of δ upward.[22] The estimated δ from the panel regressions may also differ from the IV results because of heterogeneous treatment effects.[23]

Imperfect Persistence

In our main analysis, we found $\delta < 1$, which implies that our 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 as described above. We run the regressions for both the full sample of 189 countries and for the smaller sample of 56 countries included in the main IV analysis.

Table 2 presents the results. The point estimates suggest a low degree of persistence over the period 1800–2018. 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 in the IV regressions indicates that the primary analysis may overestimate the persistence of 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.

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 facilitate 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). Glaeser et al. (2004) demonstrate low persistence of institutions between 1960 and 2000 in a cross-sectional setting.

Constant Persistence

We now investigate the assumption that δ is constant over time, which is necessary for our bias correction method. To do so, we run rolling panel data regressions with a 50-year window. For each 50-year period starting in the years between 1850 and 1963, we run a regression based on equation (28).[24] 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 0.32 and 0.76, with a mean of 0.57. The standard deviation of the coefficients is 0.08, and the figure does not reveal any obvious time trends in the estimate of δ. This stability of the estimated

---

[22]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 which are omitted variables.

[23]This is true for two reasons. First, without the instrument, we are no longer estimating effects just for the compliers. Second, in some specifications, we include a larger set of countries than in the IV regressions.

[24]Figure A.1 shows that the results are robust when using data from the 48 countries out of *all* countries in the Polity5 database for which data on Constraints on the Executive exists for at least 75 percent of the years in the period 1850–2018. Figure A.2 shows the results are also robust to the inclusion of all the 189 countries in the Polity5 database for which data on Constraints on the Executive exists for at least some years in the period 1850–2018.

Table 2: Panel Data Estimates of Persistence

	Constraint on the Executive					
	Main Sample (56 Countries)			Full Sample (187 Countries)		
	1-Year	5-Year	10-Year	1-Year	5-Year	10-Year
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Constraint on the Executive	0.926***	0.711***	0.598***	0.936***	0.745***	0.618***
	(0.0128)	(0.0489)	(0.0668)	(0.00646)	(0.0229)	(0.0292)
Number of Observations	6,012	1,222	598	16,352	3,316	1,642
Number of Countries	56	56	56	189	189	189
Adjusted R^2	0.899	0.648	0.555	0.917	0.707	0.606
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 level of constraints 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 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 calculated with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

persistence coefficient suggests that our assumption of a constant δ is a reasonable approximation. The estimate is always significantly below one, which reinforces our finding that the long-run effect is smaller than the conventional 2SLS coefficient estimate.

Linear Persistence

Finally, we use the panel dataset to examine the assumption that 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, after 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.

To estimate the relationship between Constraints on the Executive and its lagged value non-parametrically, we first run separate regressions of $Inst_{i,t}$ and $Inst_{i,t-1}$ on time and country fixed effects using a period length of five years.[25] We then capture the residuals from each regression and run a 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 (28). We also use the two sets of results to construct a flexible estimate of δ using kernel-weighted local-mean smoothing.[26]

The results are presented in Figure 3. Importantly, the non-parametric and linear regression lines are generally very close to one another. The non-parametric fit only deviates notably from

---

[25]The conclusion of the linearity assessment is robust to the use of alternative lag lengths. Appendix Table A.3 shows that the non-parametric fit remains approximately linear when using a 10-year data.

[26]We use an Epanechnikov kernel and a rule-of-thumb bandwith as defined in Stata 14's lpoly command.
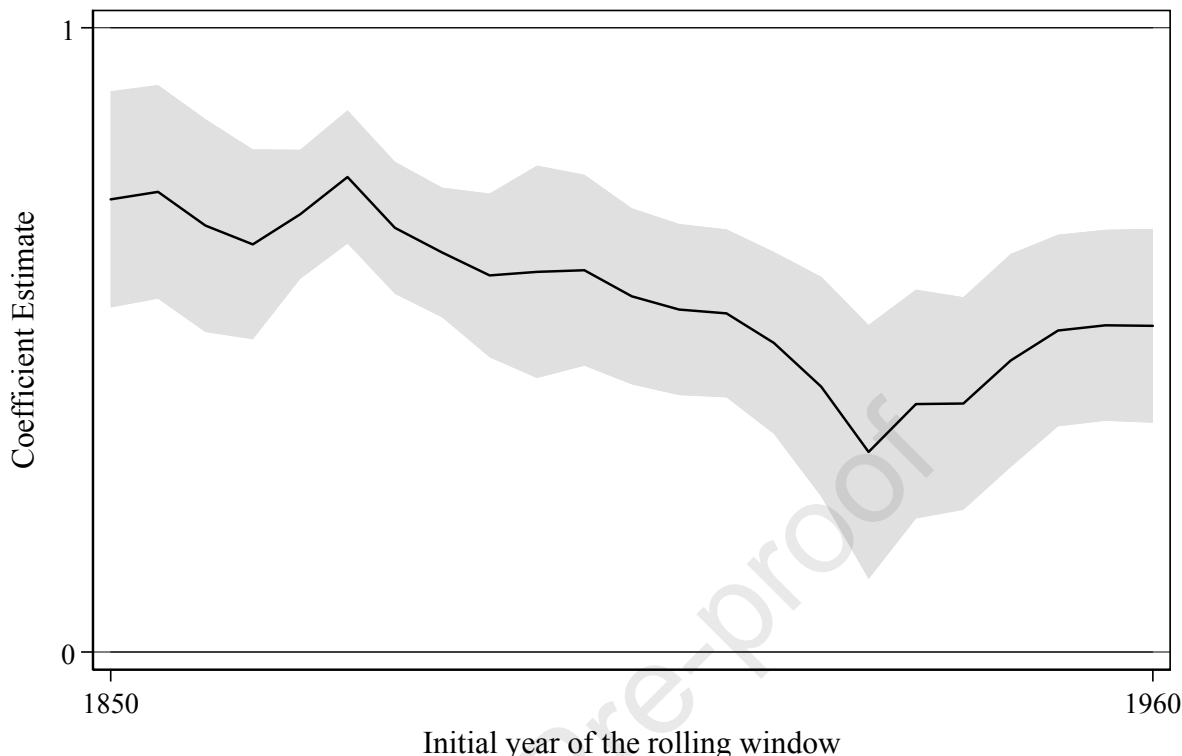
Figure 2: This figure depicts the coefficient from five-year panel regressions of Constraints on the Executive on its lagged value in over period 1850–2018 with a 50-year regression window and a step size of five years, estimated with OLS. The sample is restricted to those 21 countries, out of the sample of 56 countries from the main analysis, for which information on Constraints on the Executive exists in the Polity5 database for at least 75 percent of the years in the period 1850–2018. The regressions account for country and year fixed effects. Robust standard errors are used for the calculation of the confidence band.

the linear fit in the sparse extremes of the Constraints on the Executive index.[27] The similarity between the linear and the non-parametric fit suggests that linearity in past levels of Constraints on the Executive is a reasonable assumption.[28]

3.3 Contemporaneous Effects

Our primary goal in the empirical analysis is to estimate the long-run effect of institutions on income per capita, η. In this section, we take a brief detour to discuss bounding estimates of β_1 using the methodology developed by Conley et al. (2012). Our framework attaches a structural interpretation to the violation of the exclusion restriction. Priors over the size of the violation are inputs into the Conley et al. (2012) method. In particular, our method stresses the importance of the size of the indirect channel and the strength of the first stage in determining the bias in the regression. Neither of these forces has received attention in the existing literature when considering

---

[27]Furthermore, fitting a linear regression with a quadratic specification reveals that the second-order term is very close to zero (-0.001) and insignificant ($p = 0.921$), again indicating that a linear fit is appropriate.

[28]We also examine the non-linearity of the relationship using the data from the main analysis with OLS regressions (see Figure A.4 in the appendix). We again find that the linear relationship is a reasonable assumption.
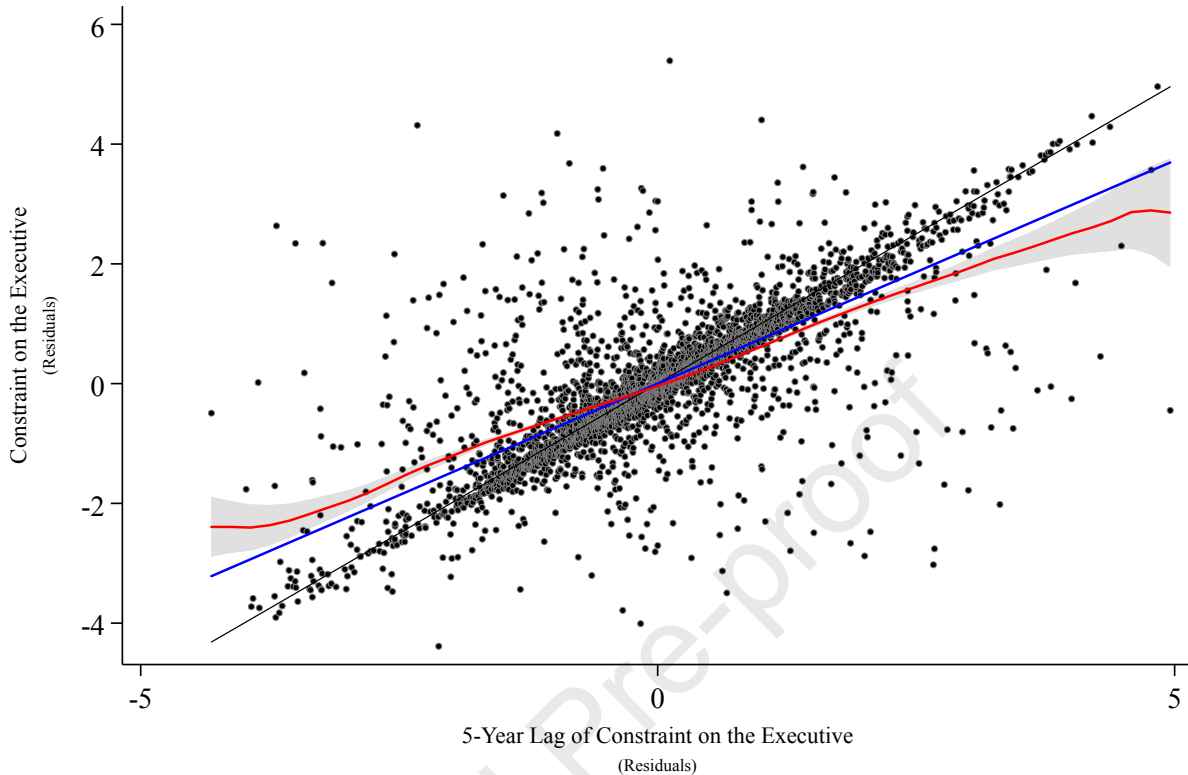
Figure 3: Linear and flexible fits of δ using a 5-year lagged panel data model accounting for country and year fixed effects over the period 1800–2018. The black line represents $\delta = 1$. The blue line represents the fit from a linear regression. The red line represents a flexible fit from a kernel-weighted local-mean smoothing. The shaded area represents the 95 percent confidence bounds of the flexible fit.

difficulties in estimating contemporaneous effects. Our goal is to highlight these forces and their importance for estimating contemporaneous effects, rather than generating precise estimates of β_1.

Using our notation, the Conley et al. (2012) method allows researchers to learn something about β_1, even when X_H affects Y_C through the A_C channel. In the Acemoglu et al. (2001) setting, the analogue of equation (10) is

$$\text{GDPpc}_{C,i} = \beta_1 \text{Inst}_{C,i} + \beta_2 \gamma \psi \text{SettMort}_{H,i} + \text{error}. \tag{29}$$

To implement the Conley et al. (2012) method, we need priors for the value of $\beta_2\gamma\psi$. The total effect of X_H on Y_C is given by $\eta = \beta_1\delta + \beta_2\gamma$, where $\beta_1\delta$ is the impact mediated by X_C and $\beta_2\gamma$ is the impact mediated by A_C. Implementing the method requires assumptions about the relative importance of indirect versus direct channels through which X_H affects Y_C. For the purposes of this example, suppose that the indirect channels, $\beta_2\gamma$, represents 10%-25% of the total impact of X_H. Focusing on the specification from column 3 of Table 1, this assumption combined with our point estimate for η implies that $\beta_2\gamma \in (0.0118, 0.0295)$. For priors over ψ, we again turn to our framework. In particular, the estimation of equation (26) gives plim $\hat{f}_1 = \psi\delta^{T-Q-H} = -1.37$ with a confidence interval of (-2.04, -0.70). We can estimate δ^{T-Q-H} as $\hat{a}_1^{\frac{T-Q-H}{Q}} = 0.41^{\frac{1900-1800}{60}} = 0.23$.

Applying this adjustment to the confidence interval for plim $\hat{f}_1$ yields $\psi \in (-8.869, -3.043)$. Putting everything together, we find that $\beta_2 \gamma \psi \in (-.26, -0.04)$. Following Conley et al. (2012), we estimate β_1 separately for different values of $\beta_2 \gamma \psi$ in this range. Taking the union of confidence intervals from each of these regressions, we are left with a confidence interval of $(-.26, 1.07)$ for β_1.

As noted above, our goal is to stress the method for bounding estimates of β_1, rather than obtaining precise bounds. Our bias correction method can be used to extract estimates of ψ. To fully understand the contemporaneous relationship between institutions and income per capita, future work needs to examine the strength of the indirect channels, $\beta_2 \gamma$, which have been overlooked.

4 Additional Examples

Table 3 presents a partial list of additional examples with historical instruments and contemporary endogenous variables. To be included in the table, papers must have at least one historical instrument in their main regressions. In the development literature, many papers use *geographic* instruments and contemporary endogenous variables.[29] While these papers fall within our framework, we do not include them here. In many of these cases, the forces we highlight are likely to be second order, compared to more direct violations of the exclusion restriction that come from impacts of geography on contemporary outcomes. For similar reasons, we do not include papers with measures of contemporary cultural as instruments, even when these are interpreted as reflecting history (e.g., Mauro, 1995; Hall and Jones, 1999; Alcalá and Ciccone, 2004).

As noted above, the work of Acemoglu et al. (2001) has had tremendous influence in the literature on institutions and comparative development. As a result, many subsequent papers have re-used their instrument to further understand the causes and consequences of institutional quality (e.g., Acemoglu et al., 2002, 2003) or to compare the importance of institutions and other variables, such as trade or financial development, in determining comparative development (e.g., Easterly and Levine, 2003; Rodrik et al., 2004; Acemoglu and Johnson, 2005; Acemoglu et al., 2014). In the latter case, other historical or geographic variables are generally included in the regressions. Some papers have adopted a similar identification strategy with different historical instruments (e.g., Gallego, 2010).

Regressions with historical instruments and contemporary endogenous regressors also play an important role in the literature on culture and comparative development. For example, Tabellini (2010) investigates the impact of contemporary culture on regional economic develop using historical literacy (circa 1880) and political institutions (1600-1850) as instruments. The measures of culture come from the World Values Survey, which only extends back to the 1980s. Gorodnichenko and Roland (2017) use genetics and disease prevalence around 1950 to examine how individualism affects labor productivity in 2000. The measure of culture comes from Hofstede (2001), whose surveys are only available from the early 2000s. In line with our framework, Gorodnichenko and Roland (2017) explicitly state a desire to measure historical culture and argue that instrumental variables can

---

[29]Geographic instruments are common when using location to predict trade or migration patterns (e.g., Frankel and Romer, 1999; Alesina et al., 2016), using land suitability to predict economic activity (e.g., Easterly, 2007), or using distance from Africa to predict genetic diversity (e.g., Ashraf and Galor, 2013; Arbatli et al., forthcoming).

eliminate the bias that comes from measuring culture at the wrong point in time. A contribution of our framework is to show that problems of estimation and interpretation still exist in this case.

Historical instruments and contemporary endogenous regressors have also been used to study the impact of financial development on economic outcomes. Building on the work of La Porta et al. (1997, 1998), Levine (1998, 1999) and Levine et al. (2000) use legal origin to instrument for measures of financial development in order to examine the impact of financial development on economic growth.[30] Beck et al. (2005) use these instruments to examine the impact of small businesses on economic growth, and Aghion et al. (2005) use them to look at the relationship between financial development and convergence in growth rates. Djankov et al. (2003) use legal origin to instrument for the degree of formalism in the legal system and examine the impact of formalism on various legal outcomes. Following in this tradition, the literature has recently turned to alternate historical instruments in order to examine the relationship between financial development and economic performance. For example, Pascali (2016) uses the presence of a Jewish community in 1500 Italy to instrument for financial development in the early 2000s, and Heblich and Trew (2019) use the existence of English post towns in the 1500s to instrument for financial development in the early 1800s.[31]

In urban economics, the influential work of Ciccone and Hall (1996) examines the importance of agglomeration effects by using instruments such as the existence of railroads in 1860 to predict the density of economic activity in 1988. As noted by Combes et al. (2010), using historical variables as instruments has since become standard in the urban literature. Focusing on the estimation of agglomeration effects, they argue that these instruments satisfy the exclusion restriction as long as "the local drivers of high productivity today differ from those of a long-gone past" (Combes et al., 2010, p. 27). Our framework demonstrates that this condition is necessary, but not sufficient, for the exclusion restriction to be satisfied. Even if the drivers of high productivity differ across time, past density may affect contemporary productivity through channels other than contemporary density. Glaeser et al. (2015) use the existence of historical mines to instrument for entrepreneurship. Duranton and Turner (2011, 2012) use railroads in 1898 and U.S. exploration paths from 1528–1850 to predict current roads in order to determine how roads affect traffic and employment growth. These instruments have also been used in subsequent studies examining alternate dependent variables (e.g., Duranton et al., 2014; Agrawal et al., 2017). Moretti (2004) uses the establishment of land grant colleges in the second half of 1800s as an instrument for the

---

[30]It is difficult to attach a date to legal origin, as it differs widely across countries. According to Berkowitz et al. (2003), "[f]or most countries, the relevant period is the 19th century; for some it reaches into the first half of the 20th century" (p. 167). When socialist legal origin in included, the relevant time period is later. In other cases, such as England, the relevant time period is earlier.

[31]Unlike many other papers using historical instruments, Pascali (2016) explicitly discusses the importance of timing in understanding different ways that the exclusion restriction can be violated. His discussion, however, only focuses on violations that occur because of direct impacts of Jewish communities on economic outcomes. Our framework identifies another avenue through which the exclusion restriction can be violated: past financial development can affect current economic outcomes through channels other than contemporary financial development.

college education share of the workforce in second half of the 1900s, in order to study human capital externalities at the city level.[32]

Historical Values as Instruments. — Our primary framework considers the case where a historical variable (Z) is used to instrument for a contemporary regressor (X_C). A related case occurs when the historical value of an endogenous variable (X_H) is used as an instrument for the contemporary value. In this case, the interpretation of the IV coefficient is the same as in our baseline case, as long as an A_C variable is the only violation of the exclusion restriction. Since historical data are available in this case, our method is not necessary, and η can be recovered through the reduced form regressions of the outcome (Y_C) on the historical value.[33] In these cases, which are not included in Table 3, it is more common for researchers to directly address the possibility of an A_C variable, since it would be a more standard violation of the exclusion restriction. We discuss some representative examples below.

In political economy, Glennerster et al. (2013) use past ethnic fractionalization to instrument for contemporary fractionalization with public good provision as an outcome variable. Baqir (2002) and Kessler (2014) use past legislative sizes to instrument for contemporary legislative sizes in order to explain variation in government policies across US cities. Satyanath et al. (2017) use past association membership to predict current association membership in explaining support for the Nazi party in German. In urban economics, older transportation networks are often used as instruments for current networks (e.g., Baum-Snow et al., 2017) and past density is used to instrument for contemporary density (e.g., Ciccone and Hall, 1996; Combes et al., 2008). In comparative development, Spolaore and Wacziarg (2009, 2016) investigate the impact of genetic distance between countries on income per capita and violence, while using past genetic distance as an instrument for contemporary distance. In the immigration literature, it is common to use existing population distributions to construct instruments for contemporary migration, essentially using past immigration to instrument for current migration (e.g., Card, 2001; Saiz, 2007; Ager and Brückner, 2013).

5 Practical Implications for Applied Research

In this section, we discuss the practical implications of our findings for applied research. We focus on estimating long-run effects (η). As described above, researchers interested in estimating contemporaneous relationships can use the 'plausibly exogenous' instruments framework of Conley et al. (2012) to formally bound β_1. In this case, our framework provides a structural interpretation of the violation of the exclusion restriction, which may be useful in generating priors about the magnitude of the violation.

When estimating long-run effects, the issue we raise is a type of non-classical measurement error. The instrument, Z, would be valid if the historical value of the endogenous variable, X_H,

---

[32]Relatedly, West and Woessmann (2010) use the catholic share of the population in 1900 to predict the number of private schools in 2003 in order to estimate the impacts of school competition on educational outcomes across countries.

[33]See appendix section A.1.5 for formal results.

Table 3: Partial list of relevant studies

Citation	Journal	Instrument(s)	Independent Variable(s)	Dependent Variable(s)
Ciccone and Hall (1996)	AER	Railroads (RR) 1860	Employment Density 1988	Labor Productivity 1988
Levine (1998)	JMCB	Legal Origin (LO)	Financial Development (FD) 1976–1993	GDP per capita (GDPpc) Growth 1976–1993
Levine (1999)	JFI	LO	FD 1960–1989	GDPpc Growth 1960–1989
Levine et al. (2000)	JME	LO	FD 1960–1995	GDPpc Growth 1960–1995
Acemoglu et al. (2001)	AER	Settler Mortality 1600s–1800s (SM)	Protection of Property Rights (PPR) ~1990	GDPpc 1995
Acemoglu et al. (2002)	QJE	SM	PPR ~1990	GDPpc 1995
Acemoglu et al. (2003)	JME	SM	PPR 1950–1970; Macro Policy 1970–1997	Macro Policy 1970–1997; GDP Volatility 1970–1997
Easterly and Levine (2003)	JME	SM	Governance Index ~1998	GDPpc 1995
Djankov et al. (2003)	QJE	LO	Legal Formalism ~2000	Contracting Institutions (CI) ~2000
Moretti (2004)	J. Econometrics	Land Grant College ~1850–1890	College Graduate Share 1980–1990	Wages 1980–1990
Rodrik et al. (2004)	JOEG	SM	Governance Index ~2000	GDPpc 1995
Acemoglu and Johnson (2005)	JPE	SM; LO; Population Density (PD) 1500	PPR ~1995; CI ~ 2000	GDPpc 1995; FD ~1998
Aghion et al. (2005)	QJE	LO	FD 1960–1995	GDPpc Growth 1960–1995
Beck et al. (2005)	JOEG	LO	Small Business Share ~1990	GDPpc/Poverty/Inequality Growth 1990–2000
Gallego (2010)	ReSTAT	SM; PD 1500; Pre-colonization cultures	Democracy/Decentralization 1900 (1985–1995)	Education (Edu) 1900 (1985–1995)
Tabellini (2010)	JEEA	Literacy ~1880; PPR 1600–1850	Values 1990s	GDPpc 1995–2000
West and Woessmann (2010)	EJ	Catholic Population 1900	Private Schools 2003	Edu 2003
Duranton and Turner (2011)	AER	Exploration Routes (ER) 1835–1850; RR 1898	Highways 1983–2003	Travel Distance 1995–2001
Duranton and Turner (2012)	RES	ER 1528–1850; RR 1898	Highways 1983	Employment Growth 1983–2003
Acemoglu et al. (2014)	ARE	SM; PD 1500; Protestant missionaries 1920s	Rule of Law 2005; Edu 2005	GDPpc 2005
Duranton et al. (2014)	RES	ER 1528–1850; RR 1898	Highways 2005	Propensity to Export 2007
Glaeser et al. (2015)	ReSTAT	Mines 1900	Entrepreneurship 1982	Employment Growth 1982–2002
Pascali (2016)	ReSTAT	Jewish Population ~1500	FD ~2000	GDPpc ~2000
Agrawal et al. (2017)	ReSTAT	ER 1528–1850; RR 1898	Highways 1983	Patent Growth 1983–1988
Gorodnichenko and Roland (2017)	ReSTAT	Disease Prevalence ~1950	Values ~2000	Labor Productivity 2000
Heblich and Trew (2019)	JEEA	Post Town 1500s	FD 1817	Man. Employment Growth 1817–1881

This table presents a partial list of studies that include regressions with historical instruments and contemporary endogenous regressors. The table lists the instruments, endogenous variables, and outcomes used in these regressions. Dates are often approximate due to multiple measures of the same variable or differences in the measurement year across observations. Often, papers use other instruments that do not fall within our framework and are not listed here. Papers often combine IV regressions with additional evidence. To be included in this table, the IV regressions must be central to the analysis, as judged by the authors of this work.

could be measured. Instead, researchers have data on the contemporary value of the endogenous variable, X_C. So, a first-best solution to the issues we raise is the collection of historical data. In that case, the long-run impact of the endogenous variable could be measured with standard instrumental variable tools.

In many cases, however, it is not feasible to collect data on X_H. In this case, researchers can use our bias correction method to estimate the magnitude of the long-run impact. When instrumenting for X_C, the relevant degree of measurement error is the inverse of persistence (δ^{-1}). A one unit increase in X_H leads to a δ increase in X_C. The conventional IV regression measure the impact of a one unit increase in X_C, which corresponds to a δ^{-1} increase in X_H. Our method uses a separate regression to estimate δ, which can then be used to correct the bias in the conventional IV regression. More specifically, our method estimates persistence over some intermediate period and then extrapolates this estimate to the entire period of interest. To do so, it is necessary to (i) observe the endogenous variable at two different points in time and (ii) make sufficient structural assumptions to extrapolate persistence over the entire period of interest. Point (i) is a matter of data availability. Section 3.2 demonstrates how to investigate point (ii) using panel data in the case that persistence is assumed to be constant and linear.

Given the strong data availability requirements and structural assumptions necessary for our method, it will not be possible to implement in all situations where there is a gap in time between a potential instrument and the measured value of the endogenous variable. In such situations, it is not possible to estimate the magnitude of the long-run effect. In almost all cases, however, there is still considerable value from investigating the existence and sign of long-run impacts. These goals can be achieved by focusing on the reduced form and first-stage regressions separately, rather than combining them into an IV estimator. The reduced form regression establishes the impact of historical events on contemporary outcomes, and the first stage provides strong evidence that the endogenous variable is at least one channel through which the historical event matters. This practice is already fairly common in the literature, and our results suggest there is nothing to be gained from adding the IV estimate, unless it is possible to implement our bias correction method.

6   Conclusion

We investigate IV research designs where there is a gap between the time when the instrument first affects the endogenous variable and the time when the endogenous variable is measured. We provide a simple theoretical framework that helps interpret these regressions. Conventional IV regressions do not consistently estimate a structural parameter of interest. We show how to augment these standard regressions to estimate the long-run effect of the endogenous variable and apply our results to examine the role of institutions in economic development, following Acemoglu et al. (2001). We also discuss cases where this correction is not possible, but our framework helps make sense of existing results.

We believe that our framework will be especially important for the literature on long-run comparative development (Spolaore and Wacziarg, 2013; Nunn, 2014; Michalopoulos and Papaioannou,

2020). By definition, studies in this field consider economic outcomes over long periods of time. A key implication of our work is that empirical and theoretical approaches cannot be fully separated in this literature. Even a very simple formal representation of long-run dynamics can greatly improve 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. 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.

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, J. ROBINSON, AND Y. THAICHAROEN (2003): "Institutional causes, macroeconomic symptoms: volatility, crises and growth," *Journal of Monetary Economics*, 50, 49–123.

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 (2002): "Reversal of fortune: Geography and institutions in the making of the modern world income distribution," *Quarterly Journal of Economics*, 1231–1294.

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.

AGER, P. AND M. BRÜCKNER (2013): "Cultural diversity and economic growth: Evidence from the US during the age of mass migration," *European Economic Review*, 64, 76–97.

AGHION, P., P. HOWITT, AND D. MAYER-FOULKES (2005): "The effect of financial development on convergence: Theory and evidence," *Quarterly Journal of Economics*, 120, 173–222.

AGRAWAL, A., A. GALASSO, AND A. OETTL (2017): "Roads and innovation," *Review of Economics and Statistics*, 99, 417–434.

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

ALCALÁ, F. AND A. CICCONE (2004): "Trade and Productivity," *Quarterly Journal of Economics*, 119, 612–645.

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*, 83, 1334–1363.

ARBATLI, C. E., Q. H. ASHRAF, O. GALOR, AND M. KLEMP (forthcoming): "Diversity and Conflict," *Econometrica*.

ASHRAF, Q. AND O. GALOR (2013): "Genetic diversity and the origins of cultural fragmentation," *American Economic Review*, 103, 528–33.

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

BAQIR, R. (2002): "Districting and government overspending," *Journal of Political Economy*, 110, 1318–1354.

BAUM-SNOW, N., L. BRANDT, J. V. HENDERSON, M. A. TURNER, AND Q. ZHANG (2017): "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 99, 435–448.

BECK, T., A. DEMIRGUC-KUNT, AND R. LEVINE (2005): "SMEs, growth, and poverty: cross-country evidence," *Journal of Economic Growth*, 10, 199–229.

BENHABIB, J., A. CORVALAN, AND M. M. SPIEGEL (2011): "Reestablishing the income-democracy nexus," *NBER Working Paper*.

BERKOWITZ, D., K. PISTOR, AND J.-F. RICHARD (2003): "Economic development, legality, and the transplant effect," *European economic review*, 47, 165–195.

CARD, D. (2001): "Immigrant inflows, native outflows, and the local labor market impacts of higher immigration," *Journal of Labor Economics*, 19, 22–64.

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.

CICCONE, A. AND R. E. HALL (1996): "Productivity and the density of economic activity," *American Economic Review*, 86, 54.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2008): "Spatial wage disparities: Sorting matters!" *Journal of Urban Economics*, 63, 723–742.

COMBES, P.-P., G. DURANTON, L. GOBILLON, AND S. ROUX (2010): "Estimating agglomeration economies with history, geology, and worker effects," in *Agglomeration economics*, University of Chicago Press, 15–66.

CONLEY, T. G., C. B. HANSEN, AND P. E. ROSSI (2012): "Plausibly exogenous," *Review of Economics and Statistics*, 94, 260–272.

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.

DIPPEL, C., R. GOLD, S. HEBLICH, AND R. PINTO (2017): "Instrumental variables and causal mechanisms: Unpacking the effect of trade on workers and voters," *NBER Working Paper*.

DJANKOV, S., R. LA PORTA, F. LOPEZ-DE SILANES, AND A. SHLEIFER (2003): "Courts," *Quarterly Journal of Economics*, 453–517.

DURANTON, G., P. M. MORROW, AND M. A. TURNER (2014): "Roads and Trade: Evidence from the US," *Review of Economic Studies*, 81, 681–724.

DURANTON, G. AND M. A. TURNER (2011): "The fundamental law of road congestion: Evidence from US cities," *American Economic Review*, 101, 2616–52.

DURANTON, G. AND M. A. TURNER (2012): "Urban growth and transportation," *Review of Economic Studies*, 79, 1407–1440.

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.

EASTERLY, W. AND R. LEVINE (2016): "The European origins of economic development," *Journal of Economic Growth*, 21, 225–257.

FRANKEL, J. A. AND D. ROMER (1999): "Does trade cause growth?" *American Economic Review*, 379–399.

GALLEGO, F. A. (2010): "Historical origins of schooling: The role of democracy and political decentralization," *Review of Economics and Statistics*, 92, 228–243.

GLAESER, E. L., S. P. KERR, AND W. R. KERR (2015): "Entrepreneurship and urban growth: An empirical assessment with historical mines," *Review of Economics and Statistics*, 97, 498–520.

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.

GLENNERSTER, R., E. MIGUEL, AND A. D. ROTHENBERG (2013): "Collective action in diverse Sierra Leone communities," *Economic Journal*, 123, 285–316.

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

HALL, R. E. AND C. JONES (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 114, 83–116.

HEBLICH, S. AND A. TREW (2019): "Banking and industrialization," *Journal of the European Economic Association*, 17, 1753–1796.

HEID, B., J. LANGER, AND M. LARCH (2012): "Income and democracy: Evidence from system GMM estimates," *Economics Letters*, 116, 166–169.

HOFSTEDE, G. (2001): *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*, Sage publications.

23

KESSLER, A. S. (2014): "Communication in federal politics: Universalism, policy uniformity, and the optimal allocation of fiscal authority," *Journal of Political Economy*, 122, 766–805.

LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1997): "Legal determinants of external finance," *Journal of Finance*, 52, 1131–1150.

LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1998): "Law and finance," *Journal of Political Economy*, 106, 1113–1155.

LEVINE, R. (1998): "The legal environment, banks, and long-run economic growth," *Journal of Money, Credit and Banking*, 596–613.

LEVINE, R. (1999): "Law, finance, and economic growth," *Journal of Financial Intermediation*, 8, 8–35.

LEVINE, R., N. LOAYZA, AND T. BECK (2000): "Financial intermediation and growth: Causality and causes," *Journal of Monetary Economics*, 46, 31–77.

MARSHALL, M. G. AND T. R. GURR (2020): "Polity5: Political Regime Characteristics and Transitions, 1800–2013: Dataset Users' Manual," *Working Paper*.

MAURO, P. (1995): "Corruption and growth," *Quarterly Journal of Economics*, 110, 681–712.

MICHALOPOULOS, S. AND E. PAPAIOANNOU (2020): "Historical legacies and African development," *Journal of Economic Literature*, 58, 53–128.

MORETTI, E. (2004): "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 121, 175–212.

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

PASCALI, L. (2016): "Banks and development: Jewish communities in the Italian Renaissance and current economic performance," *Review of Economics and Statistics*, 98, 140–158.

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.

SAIZ, A. (2007): "Immigration and housing rents in American cities," *Journal of Urban Economics*, 61, 345–371.

SATYANATH, S., N. VOIGTLÄNDER, AND H.-J. VOTH (2017): "Bowling for fascism: Social capital and the rise of the Nazi Party," *Journal of Political Economy*, 125, 478–526.

SPOLAORE, E. AND R. WACZIARG (2009): "The diffusion of development," *Quarterly Journal of Economics*, 124, 469–529.

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

SPOLAORE, E. AND R. WACZIARG (2016): "War and relatedness," *Review of Economics and Statistics*, 98, 925–939.

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

WEST, M. R. AND L. WOESSMANN (2010): "Every Catholic child in a Catholic school: historical resistance to state schooling, contemporary private competition and student achievement across countries," *Economic Journal*, 120, 229–255.

A   Online Appendix

A.1   Further Analytic Results

A.1.1   No Alternative Channels

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

$$
\begin{align}
X_{H,i} &= \psi Z_{H,i} + \varepsilon_{X_H,i} \tag{A.1}\\
X_{C,i} &= \delta X_{H,i} + \varepsilon_{X_C,i} \tag{A.2}\\
Y_{C,i} &= \beta X_{C,i} + \varepsilon_{Y,i}. \tag{A.3}
\end{align}
$$

In this set-up, $Y_{C,i} = \beta\delta X_{H,i} + (\varepsilon_{Y,i} + \beta\varepsilon_{X_C,i})$. So, $\eta \equiv \frac{\partial Y}{\partial X_H} = \delta\beta$. Now, we consider the regression equation

$$
Y_{C,i} = b_0 + b_1 X_{C,i} + \tilde{\varepsilon}_i. \tag{A.4}
$$

Since there is no violation of the exclusion restriction, the IV regression with $Z_{H,i}$ as an instrument yields

$$
\text{plim}\,\hat{b_1}^{IV} = \beta. \tag{A.5}
$$

Thus, as in the more general framework, $\eta = \delta\,\text{plim}\,\hat{\beta}_1^{IV}$. So, our results for estimating η hold whether or not an A_C variable exists. See section 2 for the analysis of the baseline case.

A.1.2   Reverse Causality

The potential for reverse causality often motivates the use of IV regressions. In particular, there may be reason to believe that Y (e.g., income per capita) exerts a causal influence on X (e.g., institutions). In this subsection, we show that our core results hold when explicitly modeling this type of reverse causality. We show that our key parameter of interest, $\eta \equiv \frac{\partial Y_C}{\partial X_H}$, incorporates the impact of Y on X. As a result, the presence of reverse causality is already accounted for in our main framework.

Consider the following extension of our base framework:

$$
\begin{align}
X_{H,i} &= \psi Z_{H,i} + \varepsilon_{X_H,i} \tag{A.6}\\
Y_{H,i} &= \beta_1 X_{H,i} + \varepsilon_{Y_H,i} \tag{A.7}\\
X_{C,i} &= \tilde{\delta} X_{H,i} + \varphi Y_{H,i} + \varepsilon_{X_C,i} \tag{A.8}\\
A_{C,i} &= \gamma X_{H,i} + \varepsilon_{A_C,i} \tag{A.9}\\
Y_{C,i} &= \beta_1 X_{C,i} + \beta_2 A_i + \varepsilon_{Y_C,i} \tag{A.10}
\end{align}
$$

where $Y_{H,i}$ is the historical value of the outcome variable (e.g., historical institutions), and $\tilde{\delta}$ is the direct impact of X_H on X_C. We continue to define $\delta \equiv \frac{\partial X_C}{\partial X_H}$. All other variables are defined as in

26

section 2. As before, we assume

$$Cov(Z_H, \varepsilon_J) = 0, \quad J = X_H, Y_H, X_C, A_C, Y_C. \qquad \text{(Assumption A.1)}$$

In words, Z is randomly assigned and is therefore uncorrelated with the error terms in all of the other equations. In this updated framework, we explicitly note that X_H affects Y_H. This is likely always the case, but it does not affect any of the results in our main framework. So, we omitted the equation for notational simplicity. The key difference between this new framework and our core results is that we now allow Y_H to affect X_C. This is the sense in which we capture reverse causality. In the case of institutions, this would be the case where historical income per capita exerts a causal impact on contemporary institutions. The size of this effect is captured by φ.

In this set-up, it is clear that Y_H affects Y_C through the channel of X_C. Since X_H affects Y_H, the impact of Y_H on Y_C should be part of the cumulative impact of X_H on Y_C. Returning to our primary example, part of the reason that historical institutions affect contemporary income is that they also increase historical income per capita, which then affects contemporary institutions and contemporary income per capita. It is straightforward to show that $Y_{C,i} = (\beta_1 \tilde{\delta} + \beta_2 \gamma + \beta_1^2 \varphi) X_{H,i} + \tilde{\varepsilon}_i$, where $\tilde{\varepsilon}_i = (\varepsilon_{Y_C,i} + \beta_2 \varepsilon_{A_C,i} + \beta_1 \varphi \varepsilon_{Y_H,i})$ is a collection of error terms. Thus, we now have $\eta \equiv \frac{\partial Y_C}{X_H} = \beta_1 \tilde{\delta} + \beta_2 \gamma + \beta_1^2 \varphi$. Similarly, it is straightforward to show that $\delta \equiv \frac{\partial X_C}{\partial X_H} = \tilde{\delta} + \varphi \beta_1$. The IV estimator now yields:

$$\text{plim } \hat{b}_1^{IV} = \frac{Cov(Y_C, Z)}{Cov(X_H, Z)}, \qquad (A.11)$$

$$= \frac{\beta_1 \delta + \beta_2 \gamma + \beta_1^2 \varphi}{\delta + \beta_1 \varphi} \qquad (A.12)$$

$$= \frac{\eta}{\delta}, \qquad (A.13)$$

which is once again our standard result. This is not surprising. The new framework simply represents the impact of X_H on X_C in slightly more detail, but this does not affect any of the intuition. Thus, our core results allow for reverse causality from Y to X.

A.1.3 Omitted Variables

Our core results do not explicitly model the reason that instruments are necessary to estimate η. In this subsection, we explicitly demonstrate what the IV regressions accomplish when compared to OLS. Consider a variable, W, that exerts a causal impact on Y_C. Also, W is correlated with X_H, but unlike A_C, it is not causally affected by X_H. In the institutions example, this could be a geographic characteristic that is correlated historical institutions. We assume that W is not measured or that the econometrician is unaware of W, raising the possibility of omitted variable

bias. The system is now given by

$$X_{H,i} = \psi Z_{H,i} + \varepsilon_{X_H,i} \tag{A.14}$$

$$X_{C,i} = \delta X_{H,i} + \varepsilon_{X_C,i} \tag{A.15}$$

$$A_{C,i} = \gamma X_{H,i} + \varepsilon_{A_C,i} \tag{A.16}$$

$$Y_{C,i} = \beta_1 X_{C,i} + \beta_2 A_{C,i} + \beta_3 W_i + \varepsilon_{Y,i}. \tag{A.17}$$

As usual, we assume that Z_H would be a valid instrument for X_H when estimating the impact of X_H on Y_C, i.e.,

$$Cov(Z_H, \varepsilon_J) = 0, \quad J = X_H, X_C, A_C, Y_C. \tag{Assumption A.2}$$

For simplicity, we also assume that $Cov(W, \varepsilon_{X_C}) = 0$. Moreover,

$$Cov(W, X_H) \equiv \nu \neq 0, \tag{A.18}$$

but

$$Cov(W, Z_H) = 0. \tag{A.19}$$

Thus, the fact that Z is uncorrelated with W is the motivation for using IV. We also define $Var(X_H) \equiv \xi \neq 0$ and $Var(\varepsilon_{X_J}) = \sigma_J \neq 0, \ J = H, C$.

Now, we consider the standard regressions equation,

$$Y_{C,i} = b_0 + b_1 X_{C,i} + \tilde{\varepsilon}_i. \tag{A.20}$$

An OLS regression will pick up the association between W and X_H in the usual omitted variable fashion:

$$\text{plim} \, \hat{b}_1^{OLS} = \beta_1 + \frac{\delta\gamma\beta_2\xi + \beta_3\delta\nu}{\delta^2\xi + \sigma_C}. \tag{A.21}$$

As a result, our method of recovering η would not work in this setting. Given that Z_H is uncorrelated with W, however, the IV results are unaffected by including this extra omitted variable in the framework:

$$\text{plim} \, \hat{b}_1^{IV} = \beta_1 + \frac{\beta_2\gamma}{\delta} = \frac{\eta}{\delta}. \tag{A.22}$$

So, the IV coefficient eliminates bias caused by correlates of X_H, but OLS does not. At the same time, IV does not eliminate the effect of channels through which X_H directly affects Y_C. This is beneficial as we want to estimate the cumulative effect of all of these channels.

A.1.4 Persistence Channels

In sections 3.1, we argued that it is important not to control for channels through which X_H affects X_C. Intuitively, this is the case because δ should capture persistence through all channels in order

to be consistent with our base framework. In this subsection, we make this point more formally. 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:

$$
\begin{aligned}
X_{H,i} &= \psi Z_i + \varepsilon_{X_H,i} & \text{(A.23)} \\
A_{C,i} &= \gamma X_{H,i} + \varepsilon_{A,i} & \text{(A.24)} \\
U_i &= \tau X_{H,i} + \varepsilon_{U,i} & \text{(A.25)} \\
X_{C,i} &= \tilde{\delta} X_{H,i} + \xi U_i + \varepsilon_{X_C,i} & \text{(A.26)} \\
Y_{C,i} &= \beta_1 X_{C,i} + \beta_2 A_i + \varepsilon_{Y,i}. & \text{(A.27)}
\end{aligned}
$$

Plugging (A.24) into (A.26) yields:

$$
X_{C,i} = (\tilde{\delta} + \xi\tau)X_{H,i} + (\varepsilon_{X_C,i} + \xi\varepsilon_{U,i}). \tag{A.28}
$$

Define $\delta \equiv \tilde{\delta} + \xi\tau$. Equations (A.23), (A.24), (A.27), and (A.28) define a system that is exactly equivalent to our base case in section 2.1. As a result, our method can be used to estimate η. Importantly, this is done without controlling for U, which is not included in the updated system.

A.1.5 Historical Values as Instruments

In this subsection, we consider the case where the historical value of the endogenous variable is the instrument. The set-up is as follows:

$$
\begin{aligned}
X_{C,i} &= \delta X_{H,i} + \varepsilon_{X_C,i} & \text{(A.29)} \\
A_{C,i} &= \gamma X_{H,i} + \varepsilon_{A_C,i} & \text{(A.30)} \\
Y_{C,i} &= \beta_1 X_{C,i} + \beta_2 A_{C,i} + \varepsilon_{Y,i}. & \text{(A.31)}
\end{aligned}
$$

As before, $\eta \equiv \frac{\partial Y_C}{\partial X_H} = (\delta\beta_1 + \gamma\beta_2)$. We assume that X_H is a valid instrument for X_C, except for the existence of the A_C variable, i.e.,

$$
Cov(X_H, \varepsilon_J) = 0, \quad J = X_C, A_C, Y_C. \tag{A.32}
$$

Consider the standard regressions equation

$$
Y_{C,i} = b_0 + b_1 X_{C,i} + \tilde{\varepsilon}_i, \tag{A.33}
$$

with X_H as an instrument for X_C. This yields

$$
\text{plim } \hat{b}_1^{IV} = \beta_1 + \frac{\beta_2\gamma}{\delta} = \frac{\eta}{\delta}, \tag{A.34}
$$

29

the same result as in our baseline case. The big difference is that now X_H is observable, so our method is not necessary. Consider the reduced form regression

$$Y_{C,i} = n_0 + n_1 X_{H,i} + \tilde{\varepsilon}_{Y,i}. \tag{A.35}$$

Now, $\tilde{\varepsilon}_{Y,i} = (\varepsilon_{Y,i} + \beta_1 \varepsilon_{X_C,i} + \beta_2 \varepsilon_{A,i})$ and $Cov(X_H, \varepsilon_Y) = 0$. So, the regression will consistently estimate η. This reduced for regression will be reported in most IV studies, implying that researchers can extract estimates of the long-run effect. In the event that it is not reported, researchers can extract a point estimate by multiplying the first and second stage coefficients.[34]

---

[34]Since $Cov(X_H, \varepsilon_{X_C}) = 0$, equation (A.29) implies that the first stage consistently estimates δ.

A.2 Additional Tables

Table A.1: Summary Statistics for Cross-Country Sample (N=56)

	Average	P25	P50	P75	S.D.
	(1)	(2)	(3)	(4)	(5).
Log GDP per capita in 1990s	8.56	7.71	8.56	9.29	1.02
Constraint on Executive in 1990s	4.42	2.85	4.20	6.15	1.90
Constraint on Executive in 1960s	3.61	2.20	3	4.90	1.94
Constraint on Executive in 1900	2.30	1	1	3	2.17
Log Capped European Settler Mortality	4.47	4.23	4.36	5.50	0.94
Log Absolute Latitude	2.42	2.08	2.63	3.14	1.01

Table A.2: The Long-Run Effect of Institutions on Income Per Capita (1990s) — Using the Period 1900-1960s for the Estimation of Persistence

	Log GDP per capita in 1990s		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.272	0.113	0.144
	(0.182)	(0.216)	(0.279)
Persistency of Endogenous Variable ($\hat{a}_1$)	0.726***	0.653	0.693
	(0.183)	(0.412)	(0.452)
Conventional 2SLS Estimate ($\hat{b}_1$)	0.710***	0.407***	0.433***
	(0.177)	(0.130)	(0.135)
World Region Fixed Effects	No	Yes	Yes
Absolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.134	0.400	0.496
First Stage F-Statistic (K-P) of Conventional	19.276	4.653	4.478
First Stage F-Statistic (K-P) of Persistence	26.367	12.587	11.135
Number of Observations	56	56	56

This table reports the estimated long-run effect of constraints on the executive in 1800 on log GDP per capita in the 1990s ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1960s ($\hat{a}_1$), and the conventional 2SLS-estimate of the effect of constraints on the executive in 1990s on log GDP per capita in the 1990s ($\hat{\beta}_1$). It also reports the results of a Wald test for the null hypothesis that the persistence coefficient is equal to one. In addition, it reports the first-stage F-statistics (Kleibergen-Paap) for the conventional 2SLS regression and the persistence regression. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors calculated 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 (2019) — Using the Period 1900-1990s for the Estimation of Persistence

| | Log GDP per capita in 2019 | | |
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.391**	0.0963	0.0985
	(0.182)	(0.120)	(0.128)
Persistency of Endogenous Variable ($\hat{a}_1$)	0.751***	0.448*	0.413
	(0.164)	(0.263)	(0.258)
Conventional 2SLS Estimate ($\hat{b}_1$)	0.753***	0.602***	0.742***
	(0.166)	(0.209)	(0.232)
World Region Fixed Effects	No	Yes	Yes
Absolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.128	0.036	0.023
First Stage F-Statistic (K-P) of Conventional	27.050	4.241	3.881
First Stage F-Statistic (K-P) of Persistence	26.367	12.587	11.135
Number of Observations	56	56	56

This table reports the estimated long-run effect of constraints on the executive in 1800 on log GDP per capita in 2019 ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1990s ($\hat{a}_1$), and the conventional 2SLS-estimate of the effect of constraints on the executive in 1990s on log GDP per capita in 2019 ($\hat{\beta}_1$). It also reports the results of a Wald test for the null hypothesis that the persistence coefficient is equal to one. In addition, it reports the first-stage F-statistics (Kleibergen-Paap) for the conventional 2SLS regression and the persistence regression. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors calculated with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.

Table A.4: The Long-Run Effect of Institutions on Income Per Capita (2019) — Using the Period 1900-1960s for the Estimation of Persistence

	Log GDP per capita in 2019		
	(1)	(2)	(3)
Long-Run Effect ($\hat{\eta}_{1800}$)	0.264	0.0959	0.125
	(0.196)	(0.190)	(0.255)
Persistency of Endogenous Variable ($\hat{a}_1$)	0.726***	0.653	0.693
	(0.183)	(0.412)	(0.452)
Conventional 2SLS Estimate ($\hat{b}_1$)	0.775***	0.402***	0.432***
	(0.200)	(0.124)	(0.134)
World Region Fixed Effects	No	Yes	Yes
Absolute Latitude	No	No	Yes
Wald Test of $\hat{\delta} = 1$ p-value	0.134	0.400	0.496
First Stage F-Statistic (K-P) of Conventional	19.276	4.653	4.478
First Stage F-Statistic (K-P) of Persistence	26.367	12.587	11.135
Number of Observations	56	56	56

This table reports the estimated long-run effect of constraints on the executive in 1800 on log GDP per capita in 2019 ($\hat{\eta}_{1800}$), the estimated persistence of constraints on the executive from 1900 to the 1960s ($\hat{a}_1$), and the conventional 2SLS-estimate of the effect of constraints on the executive in 1990s on log GDP per capita in 2019 ($\hat{\beta}_1$). It also reports the results of a Wald test for the null hypothesis that the persistence coefficient is equal to one. In addition, it reports the first-stage F-statistics (Kleibergen-Paap) for the conventional 2SLS regression and the persistence regression. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level. Standard errors calculated with the delta method and using the heteroscedasticity-consistent covariance matrix are reported in parentheses.
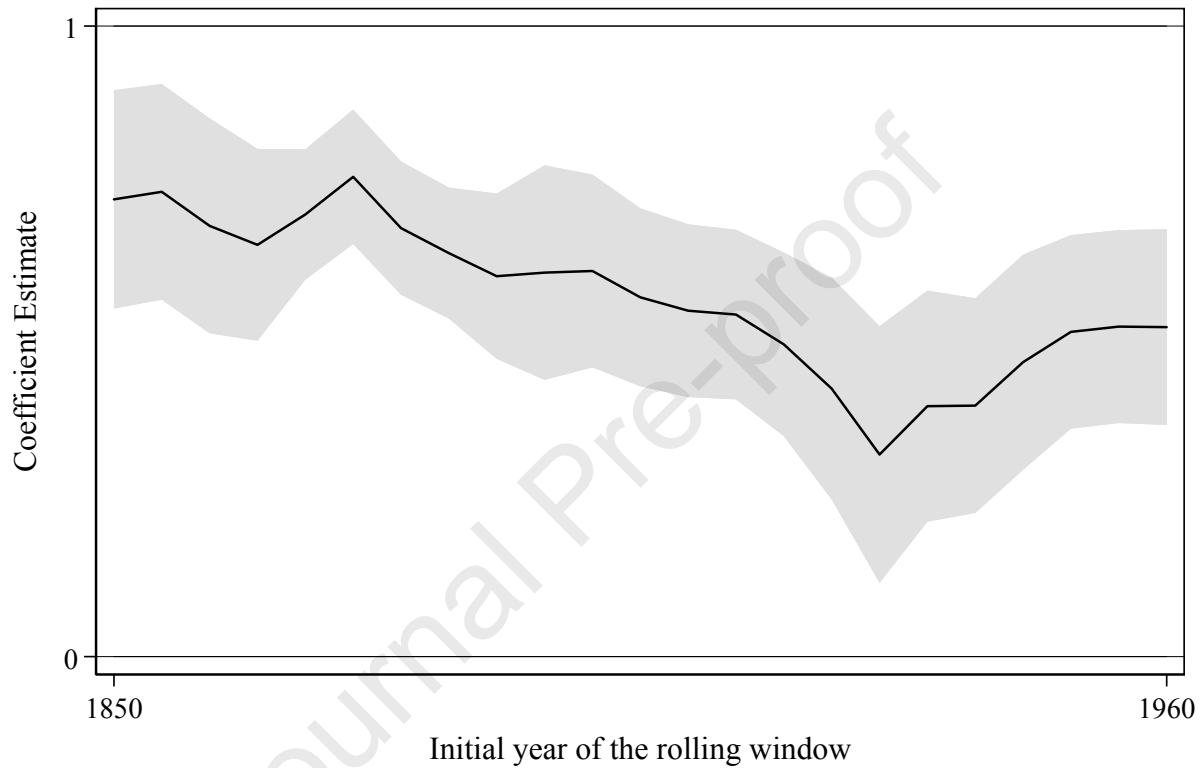
A.3 Additional Figures

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 1850–2018 with a 50-year regression window and a step size of five years, estimated with OLS. The sample is restricted to those 48 countries in the Polity5 database for which information on Constraints on the Executive exists for at least 75 percent of the years in the period 1850–2018. The regressions account for country and year fixed effects. Robust standard errors are used for the calculation of the confidence band.
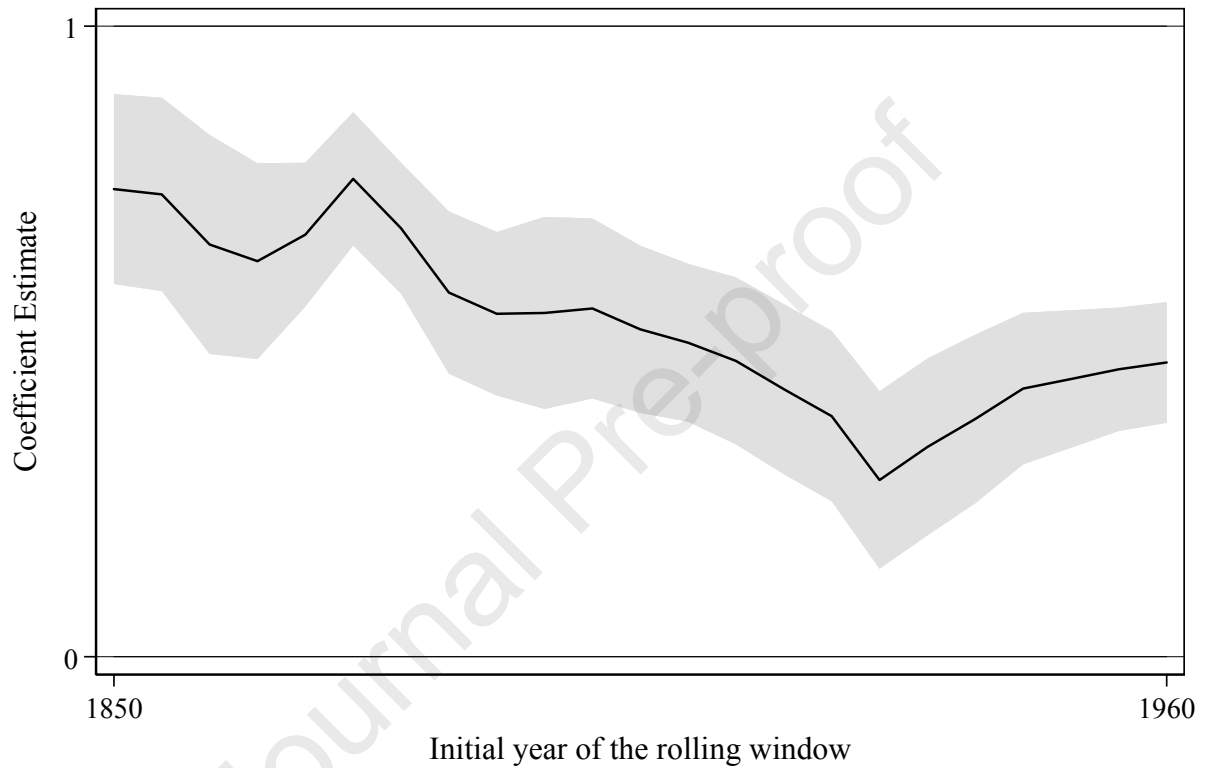
Figure A.2: This figure depicts the coefficient from five-year panel regressions of Constraints on the Executive on its five-year lagged value in over period 1850–2018 with a 50-year regression window and a step size of five years, estimated with OLS. The sample is restricted to all 189 countries in the Polity5 database for which information on Constraints on the Executive exists for at least some of the years in the period 1850–2018 and regressions can be estimated on the 10-year level. The regressions account for country and year fixed effects. Robust standard errors are used for the calculation of the confidence band.
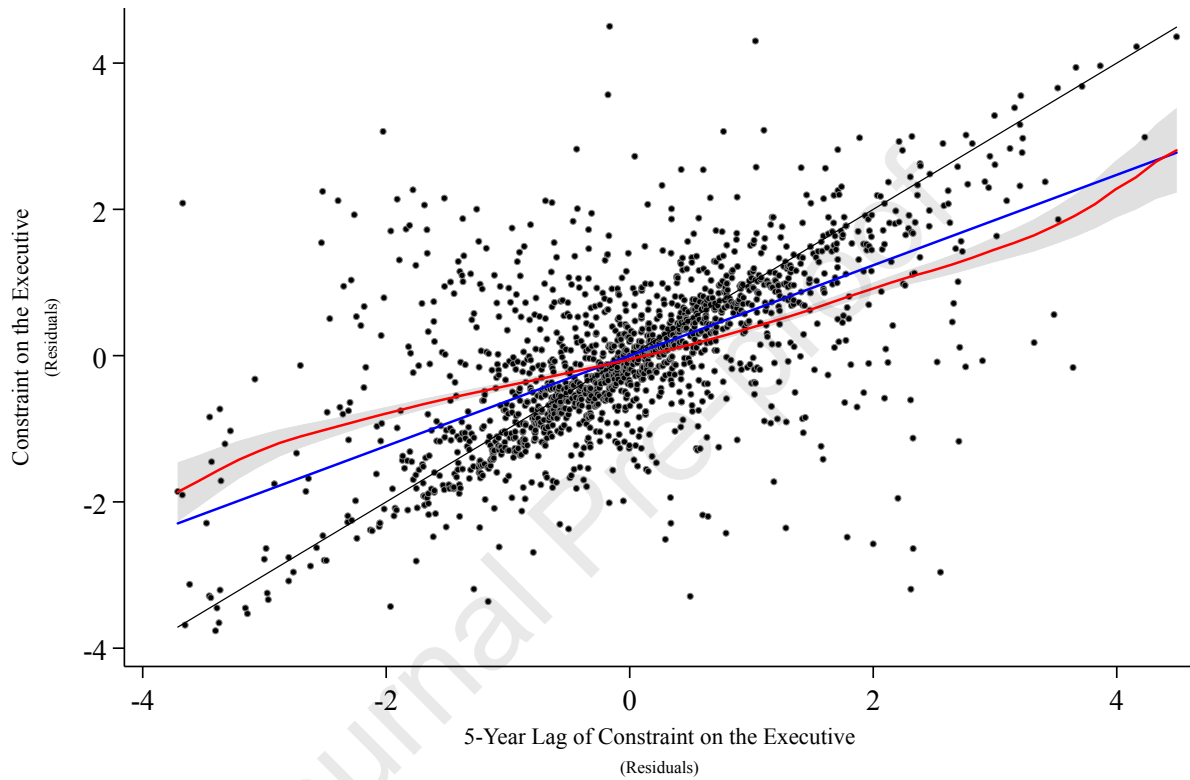
Figure A.3: Linear and flexible fits of δ using a 10-year lagged panel data model accounting for country and year fixed effects. The black line represents $\delta = 1$. The blue line represents the fit from a linear regression. The red line represents a flexible fit from a kernel-weighted local-mean smoothing. The shaded area represents the 95 percent confidence bounds of the flexible fit.

37

Figure A.4: Linear and flexible fits of δ for the cross sectional data underlying the analysis in Table 1. The black line represents $\delta = 1$. The blue line represents the fit from a linear regression. 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.