Public Disclosure Authorized

Public Disclosure Authorized

Policy Research Working Paper 9215

Leader Value Added

Assessing the Growth Contribution of Individual National Leaders

William Easterly
Steven Pennings

WORLD BANK GROUP
Development Economics
Development Research Group
April 2020

Policy Research Working Paper 9215

Abstract

Previous literature suggests that leaders matter for growth in general. This paper asks which leaders matter and develops a methodology to estimate the growth contribution of individual leaders and calculate its precision. The findings show that few leaders have statistically significant contributions; it is difficult to know who is good for growth and who is not. The paper also finds that the most intuitive estimate of a leader's contribution—the average growth rate during tenure—is largely useless for measuring his or her true contribution. Consequently, many leaders with statistically significant growth effects are surprises. Moreover, leaders in non-democratic countries are no more likely to be statistically significant than leaders in democratic ones.

This paper is a product of the Development Research Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at spennings@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

Leader Value Added: Assessing the Growth Contribution of Individual National Leaders [1]

William Easterly and Steven Pennings[2]

Keywords: National Leaders, Economic Growth, Teacher Value-added

JEL: O11, O57, O43, N10

---

[1] The views expressed here are the authors' and do not necessarily reflect those of the World Bank, its executive directors, or the countries they represent. We are grateful to Hunt Allcott, Adriana Crespo, Angus Deaton, Richard Finlay, Jenny Guardado, Aart Kraay, Ross Levine, and Laura Trucco, as well as participants in the NYU Development Seminar, NYU Development Research Institute Annual Conference, NEUDC, Midwest Political Science Association Conference, World Bank DECMG internal seminar, World Bank Kuala Lumpur Seminar, ADEW, ABCDE, Berkeley Haas Political Economy Seminar, Oxford CSAE Conference, and APSA Annual Conference. Jorge Guzman provided assistance with data.

An online appendix is available at: https://sites.google.com/site/stevenpennings/EPLeadersAppendix.pdf

[2] Easterly: NYU and NBER; email: william.easterly@nyu.edu; web: http://williameasterly.org/

Pennings (corresponding author): World Bank; email: spennings@worldbank.org

web: https://sites.google.com/site/stevenpennings/

1. Introduction

Constituents arguably consider a national leader's ability to deliver economic growth to be one of his or her most important attributes. Rewarding good economic outcomes while a leader is in office is a natural consequence of a principal-agent framework, where it is hard for voters or other constituents to directly observe the leader's abilities or to assess the quality of his or her actions (Wolfers 2007). Although previous literature addresses the importance of leaders to growth in general, we do not know how to quantitatively attribute growth to particular leaders—as constituents must—and there are no estimates of the growth contributions of individual national leaders.[3] This paper aims to fill these gaps.

Perceptions of economic competence are possibly even more important for leaders in non-democracies/autocracies (which we define using Polity data, see Section 3). These leaders have fewer constraints on their power, which arguably makes economic performance more sensitive to their intentions and abilities. Although they do not face competitive elections, leaders of non-democratic countries usually rely on the support of different constituents, who must assess if the leader is a "benevolent autocrat" or a "bad emperor." Fast economic growth forms part of the narrative justifying the rule of famous leaders such as Deng Xiaoping (China), Augusto Pinochet (Chile), Lee Kuan Yew (Singapore), and Paul Kagame (Rwanda), often to an audience well beyond their borders: media, foreign powers, foreign aid donors, and other policy makers.[4]

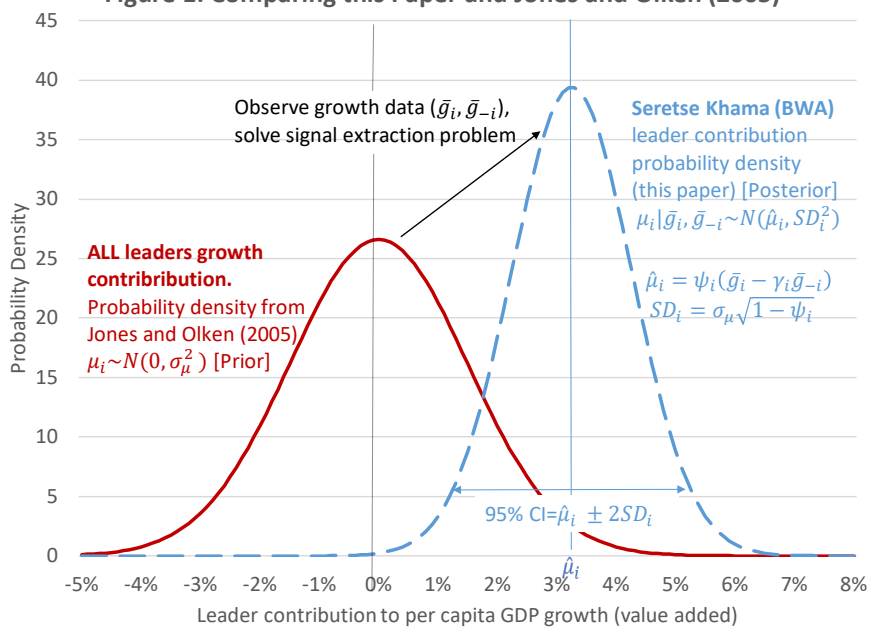
The modern leaders-and-growth literature starts with Jones and Olken (2005), who used random leader deaths in office (due to illness or accident) to investigate the effect of leaders on growth. They found that leaders have a significant effect on growth in both directions: specifically,

[3] Aruoba et al. (2019) estimate the probability that US state governors are good, based on job approval ratings.

[4] For example, Deng's *New York Times* obituary states: "In the 18 years since he became China's undisputed leader, Mr. Deng nourished an economic boom that radically improved the lives of China's 1.2 billion citizens." [link]

the standard deviation of the leader's contribution to growth is around 1.5% (for all leaders) and is statistically nonzero. As leader quality is normally distributed with mean zero in their paper, this is equivalent to Jones and Olken (2005) estimating the distribution of leader quality across all leaders, shown in the solid red line in Figure 1. Their finding is pioneering in showing that leaders matter for growth in general. But it does not tell us much about the contribution of any particular leader, which is the problem faced by constituents and international actors (and is the focus of day-to-day political and academic debates).

Figure 1: Comparing this Paper and Jones and Olken (2005)



Notes: μ_i is leader i's effect on growth, and $\hat{\mu}_i$ is its estimate. $\bar{g}_i, \bar{g}_{-i}$ represent the average growth rate residuals under leader i and other leaders in the same country (respectively). See Section 2 for further details.

In contrast, we estimate the growth contribution of every individual national leader since 1950. Even when we grant the presumption that leaders matter for growth in general, a hugely important question remains: which leaders matter? This is a signal extraction problem, where the true leader contribution is the signal we seek to extract from growth data, but it is obscured by annual country-specific noise (e.g., measurement error, good or bad luck), country-level growth trends, and global/regional shocks. Our optimal (least-squares) signal extraction method utilizes the overall

distribution of leader quality (similar to that estimated in Jones and Olken 2005) to inform the signal-to-noise ratio used to adjust the raw growth rate.[5] From a Bayesian perspective, one can consider Jones and Olken (2005) providing the prior distribution of leader quality, which we combine with the growth record during each leader's tenure (relative to other leaders in the same country) to produce a posterior distribution of leader quality for each individual.

Figure 1 shows an example of that leader estimate for Seretse Khama, the celebrated post-independence leader of Botswana (blue dashed lines), our top growth contribution leader. Khama's contribution is statistically significant because the zero is outside the central 95% part of his leader quality distribution.

Since leaders' tenures are usually short and growth rates are noisy, estimating leader contributions is not a trivial problem. For example, the intuitive unbiased estimator—the average growth rate during the leader's tenure (a leader fixed effect)—proves so inaccurate that it is better to ignore growth data entirely and allocate all leaders the same quality. There are two reasons why the average growth rate is a poor estimator of the true leader contribution. First, growth rates are usually volatile, so a high (low) average growth rate under the leader is likely to reflect good (bad) luck, rather than good (bad) policy. Second, a high (low) growth average might reflect a country effect rather than a leader effect, especially if it is similar under other leaders in the same country.

Our problem resembles the well-studied challenge of assessing the value added (VA) of schoolteachers to test scores (e.g., Kane and Staiger 2008; Chetty et al. 2014, among many others).[6] In this context, the national leader is like the teacher and noisy growth rates are like noisy

[5] "Optimal" is defined in a least-squares sense: the estimator that has the minimum expected squared deviation from the true leader effect, out of all estimators that are linear in leader's average growth rate, and the average growth rate under other leaders in the same country. See Equation 4 in Section 2.

[6] We are grateful to Hunt Allcott for suggesting the teacher value added approach.

test scores that are an imperfect measure of teacher quality. The analogous approach to the teacher VA literature is to take the average growth rate for a specific leader and shrink it towards typical leader quality. (This "shrinkage" could also increase estimated quality for leaders with below-average growth rates). Besides the application, there are three main substantive differences with the teacher VA literature. First, we calculate the precision of our estimates and hence can calculate confidence intervals and the set of statistically significant leaders. Second, we cannot identify country trends exactly as they are colinear with a sequence of leader effects and the sample is short, so it is necessary to apply a signal extraction method there as well. Finally, the set of control variables is much more limited (only regional and global shocks) given that country-level growth rates are endogenous.

Like much of the teacher VA literature, our estimates of individual leader effects are not causal as there is no exogenous variation that can be exploited for *every* individual leader. However, they are still informative about each individual leader's effect on growth, because we can address the key endogeneity issue through Monte Carlo simulation methods. The main endogeneity concern in the leaders-and-growth literature is that causality also runs from growth to leader tenure: leaders are more likely to be removed from office when growth is low (Jones and Olken 2005, Wolfers 2007). Simple regressions verify this in actual data. However, Monte Carlo simulations reveal that even if tenure is as endogenous as it is in the data, this does not affect the accuracy (root mean squared error, RMSE) or forcast bias of our least squares estimates.[7] We view this as the best answer possible to the unavoidably important question of which leaders matter, given the identification constraints inherent in producing hundreds of leader estimates.

[7] Moreover, the statistical model on which our estimates depend (Equation 1 in the next section) is in turn based on the causal estimates of the distribution of leader effects on growth from Jones and Olken (2005).

Results summary. We produce four main results. First, only roughly 45 leaders (of 650 leaders with tenures of at least 3 years with growth data for every year of their tenure) have a statistically significant growth contribution. Thus for the vast majority of leaders (93%), we cannot distinguish their growth contribution from zero using available growth data, despite using an optimal signal extraction methodology and estimating variation in underlying leader quality similar to Jones and Olken (2005). Leaders' tenure is typically either too short to reveal anything about individual leader growth contributions using growth data (low power) or too long so as to be difficult to disentangle from country effects. This difficulty in identifying individual leader effects should prompt policy makers and commentators alike to be cautious when opining that particular leaders are good or bad for growth.

Second, our estimates do confirm some famous leaders as having statistically significant positive or negative growth contributions. But they also omit others entirely and feature many surprising, forgotten, or relatively unknown leaders. Consequently, our list of statistically significant leaders looks quite different from ones featured in previous academic or policy discussions, and also from our own priors. As noted above, the leader with the largest statistically significant growth contribution is the celebrated leader Seretse Khama of Botswana (Figure 1). However, his counterpart with the most significant negative contribution is largely forgotten: General Raoul Cédras, whose ouster of Jean-Bertrand Aristide, Haiti's first democratically elected president, triggered foreign sanctions that crippled the Haitian economy.

Third, while we find evidence of higher underlying variation in leader quality in autocracies than in democracies (as do Jones and Olken 2005), leaders of autocracies do not dominate the list of significant leaders. Because autocracies have much noisier growth processes than established democracies, leaders in autocracies are not overrepresented among significant leaders. That is,

even when leaders in autocracies matter more for growth, we have trouble identifying which of them matter.

Finally, our results show that the average growth rate during a leader's tenure is virtually useless by itself to measure the leader's true effect on growth. Using the leader growth average to measure leader effects generates a strong forecast bias, a root mean squared error (RMSE) that is more than double that of the optimal methodology, and the leader growth average fails to correctly identify leaders with statistically significant contributions. It even has a worse RMSE than presuming that all leader effects are zero. The optimal method involves "shrinking" the average growth rate up or down by a factor of six toward that of the typical leader.

Contribution to the literature. Our paper addresses four literatures. First, we contribute the growth effects of specific leaders, which complements the more general literature on leaders and growth. Apart from Jones and Olken (2005), other papers estimate the general growth effect of different types of leaders (Berry and Fowler 2018; Yao and Zhang 2015), or how leaders' contributions change with education (Besley et al. 2011) or education in economics (Brown 2019).[8] Crucially, none of these papers estimates the growth contribution of individual leaders as we do.

Second, our paper contributes to a literature on electoral accountability and voter rationality (e.g., Wolfers 2007; Aruoba et al. 2019; Ashworth et al. 2017). We solve the signal extraction problem expected of rational voters who assess leaders based on economic performance. The paucity of statistically significant leaders highlights how difficult it is to determine leader quality in practice, even in hindsight and when leaders affect growth in general.

[8] Blinder and Watson (2016) study US political parties and growth. Meyersson (2016) studies the effect of coups.

Third, we speak to a literature on the role of "benevolent autocrats" in delivering growth in developing countries. According to this view, some autocrats are "growth-friendly dictators" (De Luca et al. 2015) because they have a vested interest in the whole economy and hence will produce high economic growth, an idea that goes back to Olson (1993). Glaeser et al. (2004) argue that autocratic leaders are less constrained by institutions, which makes their choices more important. "The economic success of … China most recently, has been a consequence of good-for-growth dictators, not of institutions constraining them … there was nothing pre-destined about Deng, one of the best dictators for growth, succeeding Mao, one of the worst." As such, growth under autocracy could be high or low depending on the leader, which Rodrik (2000) summarizes as the "risky gamble" of autocracies. Combined with Jones and Olken's result that autocrats matter for growth, this literature has lent credibility to a view among some aid policy makers that justifies support for repressive autocrats who deliver growth. However, such a policy requires knowledge of which leaders are good for growth. We share the recent academic enthusiasm in development to produce evidence to guide policy. Our evidence reveals very few leaders who we can confidently say have either a positive or negative growth effect.

Finally, this paper engages with a broader literature in political science and history on the effects of influential individuals versus institutions and other factors. Jones and Olken (2005) motivate their paper in part as a test of the "Great Man" theory of history, where "history is largely determined by the idiosyncratic, causative influences of certain individuals, and perhaps a very small number" (p838). While Jones and Olken (2005) showed that these "great men" do exist, we show that they are difficult to identify based on their growth records.[9]

[9] Note that being good for growth is different from being good for general welfare; many of the "Great Men" of history are almost as famous for tyranny as for economic achievements.

The rest of the paper is organized as follows: Section 2 outlines our methodology and shows that it performs well in Monte Carlo simulations, including when leader tenure is endogenous. Section 3 discusses the data and our estimates of the variability of underlying leader quality and noise. Section 4 presents our main results, including the set of significant leaders. Section 5 concludes.

2. Model and Methodology

The statistical model. We start with a simple statistical model of raw annual real per capita growth $g^{\star}_{ict}$ for leader *i* in country *c* in year *t* as in Equation 1, similar to that in Jones and Olken (2005, p840). This equation has four parts.

$$(1)\quad \underbrace{g_{ict}}_{\substack{growth\\ residual}} = \underbrace{\mu_i}_{\substack{leader\\ effect}} + \underbrace{\mu_c}_{\substack{country\\ effect}} + \underbrace{\varepsilon_{ict}}_{\substack{iid\\ noise}} \text{ where } g_{ict} \equiv \underbrace{g^{\star}_{ict}}_{\substack{raw\ growth\\ per\ capita}} - \underbrace{X\delta}_{\substack{exog.\ observables\\ (supranational)}}$$

First, it is convenient to work with the mean-zero growth residual g_{ict} after removing potentially observable exogenous supranational shocks and trends in *X* via a first-stage regression. *X* includes a region-by-year dummy variable that removes the effects of region-wide commodity price booms or busts, global growth, regional business cycles (Latin America's "lost decade"), global trend growth of ≈2%, and the fact that Sub-Saharan Africa has grown more slowly than Asia, on average.[10]

[10] We experimented with controlling for wars and country-specific commodity price shocks. This did not have much effect on the results, mostly because these shocks had poor explanatory power for growth rates. This may be due to very heterogeneous effects of commodity prices and wars. War is also problematic because it partly responds to a leader's actions.

Second, $\mu_i \sim \mathrm{N}(0, \sigma_\mu^2)$ is the effect of leader *i* on growth during his or her tenure. The goal of our paper is to estimate $\hat{\mu}_i$ for each of 1000+ leaders.[11] In contrast, Jones and Olken (2005) estimate $\sigma_\mu \approx 1.5\%$ for all leaders and show it is statistically greater than zero.[12]

Third, $\mu_c \sim \mathrm{N}(0, \sigma_c^2)$ is the country effect, which captures higher or lower trend growth in specific countries due to other factors unrelated to individual leaders such as institutions, culture, and geography.[13]

Last, $\varepsilon_{ict} \sim \mathrm{N}(0, \sigma_{\varepsilon c}^2)$ is the random *idiosyncratic* noise component of growth, with a country-specific variance $\sigma_{\varepsilon c}^2$ that is typically large (especially for developing countries). Easterly et al. (1993) show that good or bad luck swamps growth fundamentals (including leader effects), even over the medium term.

This is a model intentionally favorable to leaders. For example, it has no role for finance ministers, central bank governors, congressional representatives, or civil society.[14] The model is also not able to adjust for one-off, country-specific factors affecting growth beyond the leader's control, including some that may seem obvious to those who know the country's history.

[11] Normality of μ_i, μ_c, ε_{ict} is not required to produce estimates of $\hat{\mu}_i$ but only to calculate confidence intervals.

[12] JO normalize the standard deviation of leader quality to 1 (p. 842), such that the effect of one unit of leader quality on growth is θ, which is equivalent to σ_μ here (JO effectively test if $\sigma_\mu = 0$). Our country effect μ_c is ν_i in JO (which is nonstochastic). They allow for leader quality to be serially correlated (but we assume it is drawn independently).

[13] Country effects represent deviations from regional averages that are removed as part of $X\delta$.

[14] The only sense in which our model is unfavorable to leaders is that we regard systematically good or bad leaders as a feature of the political institutions of the country, rather than of the particular leaders. Introducing correlated leader quality in the model would be an interesting area for further work, though it substantially complicates the model.

The leader growth average (*naïve method*). A natural estimator of the leader effect $\hat{\mu}_i$ is simply the average growth rate $\bar{g}_i$ during a leader's tenure T_i (i.e., a leader fixed-effects panel regression):

(2) $\hat{\mu}_{\mathrm{i}}^{naive} \equiv \bar{g}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} g_{ict}$

The performance of the naïve method can be evaluated by a Monte Carlo simulation, which uses the real-world leader structure across countries and years but simulated growth rates based on random draws of $\{\mu_i, \mu_c, \varepsilon_{ict}\}$. This allows us to observe μ_i, the randomly drawn true leader effect, and hence evaluate the quality of the estimator. The Monte Carlo uses values of $\sigma_\mu = 1.5\%$, $\sigma_c = 1\%$, $\sigma_\varepsilon = 5\%$, similar to that estimated empirically here and also in Jones and Olken (2005).

Although conditionally unbiased,[15] the naïve leader growth average estimator performs very poorly in Monte Carlo simulations in Table 1 (panel A). Most importantly, the naïve leader growth average is not very efficient: that is, the root mean squared error (RMSE $= \sqrt{L^{-1}\sum_i(\mu_i - \hat{\mu}_i)^2}$), of the estimator is very large. This means there is typically a big gap between $\hat{\mu}_i^{naive}$ and the true leader effect μ_i.[16] A natural benchmark here is the simple alternative of just assuming $\hat{\mu}_i = 0$; that is, all leaders have the same quality (the *zero leader effect* method). The RMSE of the naïve leader growth average is 3.22%, which is more than twice as large as from just ignoring all leader effects using the zero method (1.5%). Second, the naïve leader growth average estimator is forecast biased (as in Chetty et al. 2014). A forecast *un*biased estimator means that for every 1 percentage point increase in the estimated leader effect $\hat{\mu}_i$, the true leader effect μ_i increases by 1 percentage point

[15] That is, for a fixed leader effect, the expected leader growth average is the true leader effect.

[16] Recall that the naïve method is identical to running leader fixed-effects regression, and allocating the fixed effect to the leader. If instead we also added country fixed effects (demeaning growth by country), the Monte Carlo results are similar to the naïve method in Table 1: RMSE is 3.06% (rather than 3.22%) and the forecast bias is the same.

on average. (Thus, if we estimate a regression on simulated leader data of $\mu_i = \lambda\hat{\mu}_i + e_i$, then $\hat{\lambda} = 1$.) However, for the naïve method, $\hat{\lambda}_{naive}$=0.18, meaning that when the estimated leader contribution increases by 1 percentage point, the true leader contribution only increases by 0.18 percentage points.

Why is this the case? The error of the leader growth average $\bar{g}_i - \mu_i = \frac{1}{T_i}\sum_{t=1}^{T_i} \varepsilon_{ict} + \mu_c$ includes two noise terms. $\frac{1}{T_i}\sum_{t=1}^{T_i} \varepsilon_{ict}$ is the average iid error over the tenure of the leader and is the most important. For most leaders, $\frac{1}{T_i}\sum_{t=1}^{T_i} \varepsilon_{ict}$ is a long way from zero (in either direction) because the iid noise ε_{ict} is very volatile (although it can be small for leaders with long tenures T_i). Second, the country effect μ_c contaminates all leader growth averages, even for leaders with long tenures.

Unfortunately, the country effect cannot be removed via a preliminary regression on country dummies (which is the same as subtracting the country-specific average growth rate $\bar{g}_c = \frac{1}{N_c}\sum_{t=1}^{N_c} g_{jct}$, where N_c is the number of observations per country). This is because the leader effect μ_i also contributes to $\bar{g}_c$, and so subtracting $\bar{g}_c$ would also remove part of the contribution of leader i. This is a particularly serious problem for long-tenured leaders such as Paul Biya (Cameroon) and Robert Mugabe (Zimbabwe), in whose cases subtracting the country average growth rate would remove around two-thirds of their contributions.[17]

Instead of subtracting a country dummy to remove country effects, a better approach is to subtract the average growth rate under other leaders in the same country $\bar{g}_{-ic}$, which does not

[17] The estimated country fixed effect would be $\bar{g}_c = \mu_c + \frac{1}{N_c}\sum_{t=1}^{N_c} \varepsilon_{ct} + \frac{1}{N_c}\sum_{j=1}^{L_c} T_j\mu_j$. $T_i/N_c \times \mu_i$ is part of this sum, which would be removed if subtracting $\bar{g}_c$. $T_i/N_c \approx 2/3$ for Biya (Cameroon) and Mugabe (Zimbabwe).

include the contribution of leader *i* (Equation 3). Even then, $\bar{g}_{-ic}$ is a noisy measure of the country effect because we only have around 50 observations per country, on average.[18]

(3) $\bar{g}_{-ic} \equiv \frac{1}{N_c - T_i} \sum_{j \neq i} g_{jcs}$

Table 1: Monte Carlo Estimates (mean across 500 reps)

	True: sd(leader)=1.5%; sd(country)=1%; sd(iid)=5%						
	Performance		Shrinkage Coeff		Estimated Variance components		
	Root Mean Sqaured Error (smaller is better)	Forecast Bias (λ≠1)	ψ (leader)	γ (country)	σ(leader)	σ(iid noise)	σ(country)
	Panel (A): Simple Estimators of Leader Effects						
A1) Naïve leader growth ave	3.22%	0.18	-	-	-	-	-
A2) Zero leader effect	1.50%	-					
	Panel (B): Least Squares Leader Estimates						
B1) When σ are known	1.28%	0.99	0.28	0.57	-	-	-
B2) When σ are unknown	1.28%	1.03	0.27	0.56	1.48%	5.00%	0.99%
Panel (C): Least Squares Leader Estimates (with Endogenous Leader Transitions)							
C1) When σ are known	1.28%	0.99	0.27	0.61	-	-	-
C2) When σ are unknown	1.29%	1.04	0.26	0.59	1.52%	5.19%	0.99%

Note: Table presents Monte Carlo estimates (500 reps) of variance components, the shrinkage coefficients used to construct least squares leader estimates and measures of the performance of these estimates. In Panels A and B, the actual country X leader tenure structure is used, but in Panel C, the leader leaves office with probability of 1/6-0.5growth (an average tenure of about 6 years). The growth rate of leader, country and noise are drawn from normal distributions as in Equation 1. A successful method is forecast unbiased (λ=1),has a lowest root mean squared error (RMSE), and uncovers the "true" parameter of the leader effect of 1.5%, CE of 1% and iid SE of 5%. Panel A uses simple estimators of leader effects. Panels B and C report estimates using the least squares method (Equation 5 and 6). The reported shrinkage coefficients ψ and γ used to produce the leader estimates are the means across leaders and replications. Estimated variance components are calculated using Equations 9-11.

The least-squares (LS) leader estimator. Starting with the two building blocks above—the average growth under leader *i* ($\bar{g}_i$) and the average growth rate under other leaders ($\bar{g}_{-ic}$, as a measure of the country effect)—we next try to produce a better estimator which minimizes the (squared) gap $\mu_i - \hat{\mu}_i$ by reweighting $\bar{g}_i$ and $\bar{g}_{-ic}$ according to their signal-to-noise ratio: the least squares (LS) estimator (Equation 4). This estimator is parsimonious and has simple closed forms for the signal-to-noise ratio weights on $\bar{g}_i$ and $\bar{g}_{-ic}$ (though of course, more general estimators are

[18] $\bar{g}_{-ic} = \mu_c + \frac{1}{N_c - T_i} \sum_{t=1}^{N_c - T_i} \varepsilon_{ict} + \frac{1}{N_c - T_i} \sum_{j \neq i}^{L_c} T_j \mu_j$, where the last two terms are the noise.

possible as future extensions).[19] We rearrange $\hat{\mu}_i = \beta_1 \bar{g}_i + \beta_2 \bar{g}_{-ic}$ in a more intuitive form as $\hat{\mu}_i = \psi_i(\bar{g}_{ic} - \gamma_i \bar{g}_{-ic})$, given $\bar{g}_{-ic}$ is used to subtract country effects.

(4) $\min_{\psi,\gamma} E\left[\mu_i - \hat{\mu}_i^{LS}\right]^2$ where $\hat{\mu}_i^{LS} = \psi_i(\bar{g}_i - \gamma_i \bar{g}_{-ic})$

There are two parameters to estimate in Equation 4. First, the optimal weight $\gamma_i \in [0,1]$ (given by Equation 5) reflects our ability to control for country effects using the other-leader growth average $\bar{g}_{-ic}$. This is easier ($\hat{\gamma}$ close to 1) when country effects vary greatly (high σ_c^2), and hence a high other-leader growth average $\bar{g}_{-ic}$ probably reflects a high country effect μ_c. It will be small ($\hat{\gamma}_i$ close to 0) when other factors drive variation in $\bar{g}_{-ic}$, such as (i) average iid noise (increasing with σ_ε^2, and averaging out with combined tenure of other leaders $N_c - T_i$), or (ii) variation in the quality of other leaders (increasing with σ_μ^2 but evening out with higher L_{-i}, the number of other leaders in the same country). Therefore, we can think of Equation 5 as related to the ratio of signal to noise (to be exact, the ratio of signal to signal + noise).

(5) $\gamma_i = \frac{Cov(\bar{g}_i \bar{g}_{-ic})}{Var(\bar{g}_{-ic})} = \frac{\sigma_c^2}{\sigma_c^2 + \frac{\sigma_\varepsilon^2}{N_c - T_i} + \frac{\sigma_\mu^2}{L_{-ict}}}$

Second, the optimal weight $\psi_i \in [0,1]$ on the *adjusted leader growth average* ($\bar{g}_i - \hat{\gamma}_i \bar{g}_{-ic}$) is given by Equation 6 and measures the extent to which $\bar{g}_i - \gamma_i \bar{g}_{-ic}$ reflects true leader quality. This term is known as the *shrinkage factor* because it shrinks the leader estimate $\hat{\mu}_i^{LS}$ toward zero when $\bar{g}_i - \gamma_i \bar{g}_{-ic}$ is a noisy measure of leader quality (small ψ_i). This occurs when (i) year-to-year growth is very noisy (high σ_ε^2) and the leader has a short tenure (small T_i), so a high or low $\bar{g}_i$ might be just good or bad luck; (ii) country effects vary a lot (high σ_c^2) and we cannot control

[19] For example, one could break down $\bar{g}_{-ic}$ into several terms for different other leaders in the same country $\bar{g}_{j_1 c} \ldots \bar{g}_{j_L c}$ $\forall j \neq i$, and weight each according to its signal-to-noise ratio.

for them using the other-leader growth average (γ_i close to 0). Again, Equation 6 measures the ratio of signal to noise.

$$(6)\ \psi_i = \frac{Cov(\mu_i, \bar{g}_i - \hat{\gamma}_i \bar{g}_{-ic})}{Var(\bar{g}_i - \hat{\gamma}_i \bar{g}_{-ic})} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_c^2(1-\hat{\gamma}_i) + \frac{\sigma_e^2}{T_i}}$$

Monte Carlo simulations confirm that the performance of the LS leader estimator is far superior to the naïve leader growth average (assuming $\{\sigma_\mu,\ \sigma_c,\ \sigma_\varepsilon\}$ are known; Table 1, row B1). Specifically, the RMSE of the LS leader estimator is less than half that of the naïve leader growth average method, which is unsurprising given our estimator minimizes errors by construction. Unlike the naïve method, the LS leader estimates are also forecast unbiased (λ=1). The LS leader estimates also outperform the zero leader effect method in terms of RMSE, though it is interesting that the gains are not that large (1.28% vs 1.5%). This illustrates the difficulty of estimating leader effects, even with an optimal methodology and known data-generating process. The average country adjustment is $\hat{\gamma}_i$=0.57 (across leaders), reflecting the fact that $\bar{g}_{-ic}$ is a noisy measure of country effects. Moreover, the average leader shrinkage factor is $\hat{\psi}_i$ =0.28 (across leaders), showing that the [adjusted] leader growth average is a very poor measure of leader quality and needs to be shrunk substantially toward zero.

The final panel of Table 1 assumes that leader transitions are endogenous (depending on growth rates), rather than exogenous as in the rest of Table 1. It shows that even with endogenous transitions, our LS methodology performs just as well (see the end of this section for a discussion).

Relation to teacher VA literature. Similar methods have been used quite successfully to assess the VA of schoolteachers to test scores.[20] The contexts here and in that literature are

[20] For example, Chetty et al. (2014) find that VA estimates predict changes in test scores in event studies where teachers change schools, and Kane and Staiger (2008) utilize a random assignment of teachers to show that teacher VA estimates are conditionally unbiased. These methods are not without criticism, however. For example, Rothstein

different, which means that (i) the set of exogenous covariates removed in $X\delta$ in Equation 1 is much more limited and (ii) we can noisily observe country effects and so must apply a shrinkage approach there as well. Abstracting from country effects, observables, and some other teacher-specific terms, the shrinkage factor in Equation 6 is the same as Equation 9 in Chetty et al. (2014) and Equation 5 in Kane and Staiger (2008).[21] Another important difference is that we estimate the precision of leader estimates and the set of "statistically significant" leaders (discussed next).

Precision and the set of statistically significant leaders. In addition to knowing the "best" point estimate of the leader effect, it is important to know how accurate that estimate is and whether we are confident it is positive or negative. Since we assume leader quality and growth averages are normally distributed, so are the LS leader estimates:

$$(7) \qquad \mu_i|\bar{g}_i, \bar{g}_{-ic} \sim N(\hat{\mu}_i^{LS}, \sigma_\mu^2(1-\psi_i)) \text{ where } \hat{\mu}_i^{LS} = \psi_i(\bar{g}_i - \gamma_i \bar{g}_{-ic})$$

Equation 7 says that the distribution of leader quality, conditional on observing $\bar{g}_i$ and $\bar{g}_{-ic}$, is normally distributed with a mean of the LS leader estimate $\hat{\mu}_i^{LS}$ (from Equation 4-6) and a standard deviation $\sigma_\mu\sqrt{1-\psi_i}$, where ψ_i is the shrinkage factor from Equation 5. This can be derived from the properties of multivariable normal distributions (see Online Appendix 1).[22] Alternatively, one can take a Bayesian perspective where μ_i^{LS} is both the maximum and mean of the posterior

(2010) finds that some assumptions of VA models are violated, which can lead to future teachers affecting past test scores, and that teacher VA estimates fade out quickly.

[21] Specifically, to get the equivalence, one needs to abstract from country effects ($\sigma_c^2 = 0$) and classroom effects ($\sigma_\theta^2 = 0$) and assume there is one student per class ($n = 1$). Teacher's tenure ($t-1$) and the variance of student error σ_ε^2 in Chetty et al. (2014) correspond to leader tenure T_i and the variance of annual GDP growth errors here.

[22] We thank Aart Kraay for pointing out this way to derive the least squares mean and variance.

distribution $f(\mu_i|\bar{g}_i, \bar{g}_{-ic})$, and $\sigma_\mu\sqrt{1-\psi_i}$ is its standard deviation. There is a 95% probability that the true leader effect μ_i lies within the interval:[23]

(8) 95% CI= $[\hat{\mu}_i^{LS} - 1.96 \times \sigma_\mu\sqrt{1-\psi_i}, \quad \hat{\mu}_i^{LS} + 1.96 \times \sigma_\mu\sqrt{1-\psi_i}]$

We call a leader's contribution *statistically significant* if zero is not included in the 95% CI in Equation 8. This determines leaders with a statistically significant contribution in Table 3.

For the LS leader estimates, $\text{RMSE}_{\text{LS}} = \sigma_\mu\sqrt{1-\psi_i}$, which links the accuracy estimates in the Monte Carlo in Table 1, the width of CIs, and the size of the shrinkage factor ψ_i.[24] As $\psi_i \to 0$, for example when $\sigma_e^2, \sigma_c^2 \to \infty$, then the $\text{RMSE}_{\text{LS}} \to \sigma_\mu$. Hence the RMSE_{LS} is bounded above by the RMSE of the zero-leader effect method. The shrinkage factor is also a measure of how noisy the estimates are. As $\psi_i \to 1$ (for example $\sigma_e^2, \sigma_c^2 \to 0$), the $\text{RMSE}_{\text{LS}} \to 0$.

Estimates of true population variance components. The final step before applying the model to the data is to estimate the variance components $\{\sigma_\mu, \sigma_c, \sigma_\varepsilon\}$ used to construct ψ_i and γ_i. A natural estimator of the iid error variance $\hat{\sigma}_\varepsilon^2$ is the variability of the growth residual *within* leader terms, as in Equation 9 (this method is used by Stata's *xtreg, re* and *xtreg, sa* commands; the formula provided is for balanced panels, with $T_i = T$). We estimate $\hat{\sigma}_\varepsilon^2$ country by country, which is important as many countries with famous leaders also have very noisy iid growth rates, and is feasible as $\hat{\sigma}_\varepsilon^2$ is estimated accurately with only ≈50 observations.

(9) $$\hat{\sigma}_{\varepsilon,c}^2 = \frac{1}{N_L(T-1)}\sum_{i=1}^{N_L}\sum_{t=1}^{T}(g_{ict} - \bar{g}_{ic})^2$$

The country effect is estimated based on the covariance of growth rates in the same country under different leaders as $cov(g_{ict}, g_{jcs}) = \sigma_c^2 \ \forall t \neq s, i \neq j$, implemented as:

[23] Technically, this is a 95% credible interval, which is the Bayesian analog of the classical confidence interval.

[24] E.g., Table 1, row B2, $\sigma_\mu = 1.5\%$ and $\bar{\psi}_i = 0.28$, so $\text{RMSE}_{\text{LS}} = 0.015 \times \sqrt{1-0.28} = 1.3\%$.

(10) $$\hat{\sigma}_c^2 = \frac{\sum_c \sum_{i \neq j; s \neq t} g_{ict} g_{jcs}}{\sum_c \sum_{i \neq j; s \neq t} 1_{ict} 1_{jcs}}$$

The leader effect is estimated based on the covariance of growth rates in the same country under the same leader as $cov(g_{ict}, g_{ics}) = \sigma_\mu^2 + \sigma_c^2 \ \forall t \neq s, \forall\, i, \forall c$. As this includes σ_c^2, we need to subtract that (using $\hat{\sigma}_c^2$ from Equation 10):[25]

(11) $$\hat{\sigma}_\mu^2 = \frac{\sum_{i,c} \sum_{; s \neq t} g_{ict} g_{ics}}{\sum_{i,c} \sum_{s \neq t} 1_{ict} 1_{ics}} - \hat{\sigma}_c^2$$

Monte Carlo evidence from Table 1, row 2B suggests the estimators in Equations 9-11 are extremely accurate: the average $\hat{\sigma}_\mu, \hat{\sigma}_c, \hat{\sigma}_\varepsilon$ across Monte Carlo replications is extremely close to the true values in the Monte Carlo of $\sigma_\mu = 1.5\%, \sigma_c = 1\%, \sigma_\varepsilon = 5\%$. Consequently, the performance statistics ($RMSE_\mu$ and bias $\hat{\lambda}$) and average shrinkage coefficients (ψ_i and γ_i) are very similar to those when the variance components $\{\sigma_\mu, \sigma_c, \sigma_\varepsilon\}$ are known in Table 1, row B1.

Endogeneity The key empirical challenge in estimating the effect of leaders on growth is that "leader transitions are often non-random, and may in fact be driven by underlying economic conditions" (Jones and Olken 2005, p836). While we cannot address this issue through a quasi-experimental design like Jones and Olken (2005) for hundreds of leader estimates, we can show that the most plausible kind of endogeneity in the data does not affect the performance of our empirical methodology. The first step in this approach is to estimate a linear probability model of how growth affects the probability of a leader transition in the actual leader-growth panel data:

[25] Wooldridge (2001, p260) advocates using the covariance between errors within the same unit as a measure of the individual effect. Chetty et al. (2014) and Kane and Staiger (2008) also estimate teacher effects via the covariance between average test scores in each class taught by the same teacher each year. An alternative approach used in previous versions of this paper (and in Stata's *xtreg, re* and *xtreg, sa* commands) is to first estimate the between variance (in our context, the variance across leader growth averages $\hat{\sigma}_b^2 = N_L^{-1} \sum_{i=1}^{N_L} \bar{g}_i^2$), and then subtract the iid error variance in Equation 9. However, we found that this method was upward biased ($\hat{\sigma}_\mu^2$ too large) in Monte Carlo simulations, quite noisy, and often had $\hat{\sigma}_b^2 - \hat{\sigma}_{\varepsilon,c}^2 < 0$, which was rounded to zero.

(12) $$\mathbf{1}(LeaderTransition)_{ct} = P_{trans} + \beta g_{ct} + \mu_c + e_{ct}$$

where $\mathbf{1}(LeaderTransition)_{ict}$ is a dummy variable that equals one if a change in leader occurred in year *t* in country *c* and g_{ct} is the contemporaneous growth residual. Estimation with country fixed effect (μ_c), yields an estimate of $\hat{\beta} = -0.5$ (*t-stat=-6*), verifying the endogeneity of leader transitions: a 1ppt lower growth rate increases the probability of leader transition by 0.5ppts.[26]

To test the effect of this endogeneity on our methodology, Panel C of Table 1 re-produces all of the least-squares Monte Carlo estimates when leaders change endogenously with the growth rate, as in Equation 12 with $\hat{\beta} = -0.5$.[27] Despite being very statistically significant, the potential endogeneity of leader tenure has almost no effect on any of the least squares Monte Carlo estimates. Why? This form of endogeneity leads to an upward bias in average growth rates at long tenures due to sample selection bias (leaders with higher growth rates are more likely to survive). But its quantitative size is small, and so selection bias is miniscule at the average tenure of about 6 years. Moreover, it seems that in terms of RMSE, sample selection bias is a second-order issue relative to shrinking iid error noise.[28]

3. Data and estimates of variance components

Data sources. Data on growth are taken from Penn World Tables (PWT) version 9 over 1951-2014 (the latest version at the time of writing, Feenstra et al. (2015), and data on leaders are taken from the Archigos 4.1 data set (described in Goemans et al. 2009).[29] We categorize countries as

[26] Another potential interpretation is that the turbulence of the leader transitions causes lower growth.

[27] That is, the probability of a leader change in year is 1/6 -0.5*growth (recall the growth rate is mean zero).

[28] For example, even if endogeneity is 10 times as powerful in the data ($\beta = -5$), the RMSE of the LS leader estimate only increases by 0.06ppts, which is mostly due to biased variance estimates of the leader effect (σ_μ).

[29] $g_t^\star \equiv \ln(\mathrm{Y}_t) - \ln(\mathrm{Y}_{t-1})$, where Y_t is real GDP per capita (rgdpna/pop in PWT9). If multiple leaders hold office in a year in the same country, we allocate that year to the leader in power for the most days.

established democracies (hereafter "democracies" or DEM) if they have an average Polity IV score >7.5 (Marshall et al. 2016), and as *autocracies or transition countries* (hereafter "autocracies" or AUT) otherwise. The purpose of the high polity cutoff is to exclude countries that transition in and out of democracy, since countries are categorized once as one regime or the other.[30] Combined, there are around 125 countries, 7,000 observations, and 1,100 leaders for which we have growth, leader, and polity data. Of these, 23% are established democracies. (See the Online Appendix for further details and descriptive statistics.) For our results involving individual leaders' growth estimates, we restricted the sample to leaders with tenure ≥ 3 years, and growth data available for their whole tenure, leaving a sample of approximately 650 leaders and 5,400 observations.

There are six regions (for region-by-year fixed effects), which are based on the World Bank classification, with three modifications: (i) European offshoots (the United States, Canada, Australia, and New Zealand) are added to Europe, as is common in the literature; (ii) South Asia and East Asia & Pacific are combined, due to the small number of countries in the former; and (iii) Communist transition countries geographically close to the former USSR form their own region, as they experienced a distinct set of common shocks, especially in the early 1990s.[31]

Outliers. Per capita GDP growth rates are often very volatile—for example, around wars—and a small number of extreme observations can have a large effect on estimated results. Accordingly, we drop outliers, which we define as $|g_t^\star|$>40%. This is quite conservative and only

[30] AUT includes autocracies as well as democratic-transition countries that spent much of the sample as autocracies or anocracies, even if they are now democracies. Several small countries without a polity score have been dropped.

[31] The six regions are: Sub-Saharan Africa (SSA); Middle East and North Africa (MENA); Latin America and the Caribbean (LAC); Asia (a combination of World Bank regions of East Asia & Pacific, and South Asia); Europe and European offshoots; and Communist Bloc countries close to the USSR (Albania, Bulgaria, Hungary, Mongolia, Poland, and Romania). Countries with fewer than 30 years of growth data are dropped, which removes other newly created post-Communist countries. We also exclude Kuwait in 1990-91 as it was occupied and not a separate country.

accounts for 0.2% of the sample (Online Appendix Table 2). Some of the largest outliers include Iraq during the Gulf War, Rwanda during the 1994 genocide, and Lebanon during the civil war in the 1970s-1980s; different data sets disagree dramatically as to growth rates during these periods. We also drop Liberia as it has many extreme growth outliers, and Myanmar due to irregularities in its PWT9 growth data (e.g., a three-fold increase in GDP per capita in 1970).

A major caveat is that our results are only as good as our growth data. These are not only noisy but also possibly even manipulated. For example, Martinez (2018) uses the brightness of nighttime lights as a proxy for GDP and finds that autocracies systematically overstate GDP growth. We address this data quality issue in Online Appendix 3 by calculating leader effects using two other growth data sources and find that the set of significant leaders often changes.

Estimates of variance components in the data. Table 2 shows estimates of the variance components $\{\sigma_\mu, \sigma_c, \sigma_\varepsilon\}$ used to construct shrinkage factors ψ and γ for the LS leader estimates. They are calculated using PWT9 growth data and Equations 9-11 above. In the pooled sample of all leaders, our estimate of the underlying standard deviation of leader quality is $\hat{\sigma}_\mu = 1.33\%$, which is similar to Jones and Olken's (2005) all-leaders estimate of $\hat{\sigma}_\mu = 1.5\%$. The standard deviation of the country effect is $\hat{\sigma}_c = 0.65\%$, and the standard deviation of the iid error is much larger, around $\hat{\sigma}_e$ =4.6%.

Table 2: Estimates of Variance Components in PWT9 Growth Data

1. Pooled			2. Autocracies/Transition			3. Established Democracies		
SD(leader)	SD(CE)	SD(iid)*	SD(leader)	SD(CE)	SD(iid)*	SD(leader)	SD(CE)	SD(iid)*
1.33%	0.65%	4.58%	1.35%	0.72%	4.99%	1.04%	0.42%	2.62%
(0.13%)	(0.13%)	(0.21%)	(0.15%)	(0.15%)	(0.23%)	(0.18%)	(0.12%)	(0.27%)

Notes: Estimates of variance components using the method described in the text (Equations 9-11) on PWT9 data (excluding outliers). Established Democracies are defined as countries with an average polity score>7.5 (Autocracies/Transition Countries otherwise) *SD(iid) is the across-sample SD of the iid error, though in the estimation of leader effects, we allow this to vary across countries. Standard deviations of estimates based on a country-level block bootstrap are shown in parentheses.

In constructing ψ and γ for the leader estimates, we allow $\hat{\sigma}_\mu$ and $\hat{\sigma}_c$ to vary across autocracies and democracies, and allow $\hat{\sigma}_e$ to vary by country. The motivation for allowing $\hat{\sigma}_\mu^{AUT}$ and $\hat{\sigma}_\mu^{DEM}$ to differ is that non-democratic leaders have fewer constraints on executive power; this fact arguably makes economic performance more sensitive to their intentions and abilities. This is borne out in the data: the standard deviation of leader quality in autocracies is $\hat{\sigma}_\mu^{AUT} = 1.35\%$, which is higher than leaders in established democracies ($\hat{\sigma}_\mu^{DEM} = 1.04\%$) (although they are not statistically different based on a country-level block bootstrap). Jones and Olken (2005) also found leader quality varies more in autocracies than democracies.

Democratic countries are also a more homogenous and higher-income group, which might mean $\hat{\sigma}_c^{AUT} > \hat{\sigma}_c^{DEM}$ and $\hat{\sigma}_e^{AUT} > \hat{\sigma}_e^{DEM}$ (respectively). Again, this is what we find, as $\hat{\sigma}_c^{AUT} = 0.72\%$ and $\hat{\sigma}_c^{DEM} = 0.42\%$, though they are insignificantly different (based on the same bootstrap exercise). Crucially, year-to-year growth variation (the iid error component) is higher in autocracies than democracies, with $\hat{\sigma}_e^{AUT} = 5.0\%$, almost double that of $\hat{\sigma}_e^{DEM} = 2.6\%$, and is statistically different (p-value=0%).

In sum, $\hat{\sigma}_\mu^{AUT} > \hat{\sigma}_\mu^{DEM}$ means that leader quality varies more across leaders in autocratic countries than across democratic ones, but $\hat{\sigma}_e^{AUT} > \hat{\sigma}_e^{DEM}$ and $\hat{\sigma}_c^{AUT} > \hat{\sigma}_c^{DEM}$ mean that growth is also noisier in autocracies, making leader effects more difficult to identify. These two factors are mostly offsetting in determining the fraction of statistically significant leaders.

4. Results: Which leaders have statistically significant growth effects?

Now that we have estimates of the variance components (from Table 2), we can use them to produce LS estimates of the leader effect and calculate the precision of that estimate for every national leader in our data set using Equations 5 and 6. On average, a 1 percentage point increase

in the raw leader growth average is only associated with a 0.17 percentage point increase in the LS leader effect, suggesting high average growth rates are only weakly informative about an individual leader's performance (Online Appendix Figure 1).[32]

Statistically significant leaders. We define the best (and worst) leaders for growth as any leader for whom the estimated LS leader effect is positive (or negative) and significant at the 95% level. Figure 2 plots the combination of the LS leader estimate and the size of its standard deviation for autocracies and democracies (each leader with tenure ≥ 3yrs is a dot).[33] Leaders who are significantly good for growth are in the bottom right corner and their negative counterparts are in the bottom left corner. Both have leader effects (in absolute value) more than twice as large as their error. The significant leaders are listed in Table 3. Leaders with insignificant growth contributions are in the center region.

Our first main result is that the vast majority (93.5%, 602/646) of leaders with tenures ≥ 3yrs and complete growth data have a contribution insignificantly different from zero.[34] That is, these leaders have errors too large relative to their LS leader effect. This does not cast doubt on the existence of leader effects in general, since Table 2 shows sizable variation in underlying leader quality σ_μ (as in Jones and Olken 2005). Rather, the low number of significant leaders instead exposes the difficulty of detecting which leaders represent a draw of high or low quality amid the

[32] Online Appendix Figure 2 shows that most of this difference is due to the shrinkage factor ψ.

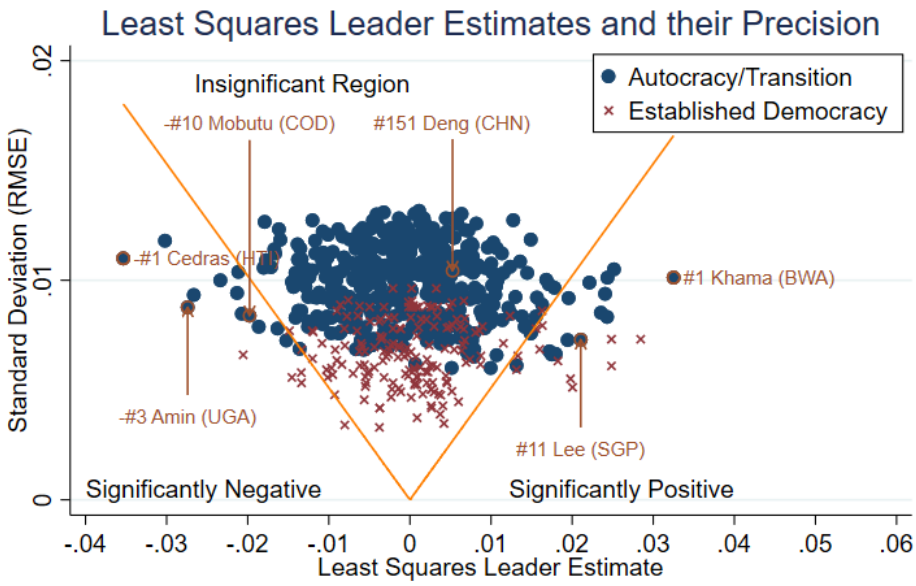
[33] We focus on leaders with tenures of 3 years or more because annual growth data introduce substantial rounding errors for leaders with shorter tenures. We also exclude leaders where we have missing growth data for some of their tenures - leaders who are in office outside 1951-2014 - as we do not observe their full growth record.

[34] Nothing guarantees that only a small fraction of leaders are significant. Monte Carlo evidence shows that if growth were less noisy (an iid SD of 1% rather than 5%), 40% of leaders would be statistically significant at the 95% level.

other factors and high noise. Even though good and bad leaders exist, we generally do not know who they are.

Our second main result is that leaders in autocracies are not overrepresented among those who are significantly good or bad for growth. Even though many leaders in autocratic countries have larger positive or negative LS leader effects, and underlying leader quality varies more in autocratic countries ($\sigma_{\mu}^{AUT} > \sigma_{\mu}^{DEM}$), it is harder to distinguish leaders in autocratic countries who are good or bad for growth from the rest due to the greater noise. This is striking in Figure 2, where the RMSE is typically 0.75-1.25% for leaders in autocratic countries but 0.5-0.9% for democratic leaders. Overall, democratic leaders account for 11 of the 26 (42%) top statistically significant leaders and 5 of the 18 (28%) significant negative-growth leaders, but they comprise only 27% of leaders with tenures ≥ 3yrs for whom we have complete growth data.

Figure 2



Notes: Orange lines mark the border of statistical significance, where 1.96*SD=abs(leadereffect) Each LS leader effect is a dot/cross. Open circles are named leaders. Sample with tenure>=3yrs and complete growth data. Leaders in SE(SW) corner are significantly positive (negative).

A review of some of the best and worst leaders for growth helps to show why our method diverges so much from the prevailing practice of giving leaders credit for the raw growth average during their tenures.

Region-year adjustments. The first adjustment in producing the LS leader estimates is to remove region-by-year effects, which reduces the magnitude of almost all leader estimates (Appendix Figure 2A). This adjustment removes around 2 percentage points from all adjusted leader average growth rates—as that is the world trend per capita growth rate—but it removes more or less for specific leaders depending on regional growth averages for the whole sample and the timing of regional business cycles (Appendix Figure 5). Averaged over time, per capita growth was around 3.5% in Asia and Communist Bloc countries, around 2.4% in Europe and offshoots, and around 1.5-1.8% in other regions (Latin American and the Caribbean, the Middle East and North Africa, and Sub-Saharan Africa).

Regional growth differences plausibly reflect factors beyond the control of leaders, such as a shared regional culture or history, trade and investment networks, and geography. As growth under the typical Asian leader was 2 percentage points higher than under the typical African leader, the Asian leader needed to achieve raw growth 2 percentage points faster than the African leader in order to be allocated the same average growth residual. But regional growth also varied over time due to common shocks and business cycles in the same region, such as the Asian financial crisis, Latin America's lost decade, and the fall of the USSR. For example, Dwight Eisenhower and Richard Nixon surprisingly appear on the negative list in part because growth under them was low relative to contemporaneous high growth in Europe.

Country effects. The residual after region-year adjustment will be compared to the same residual for other leaders of the same country. While this adjustment is not so important on average,

it is important for many individual leaders (Appendix Figure 2B). For example, the significant negative growth effect of Junichiro Koizumi (Table 3, panel B) is in part because of high growth under other Japanese leaders such as Hayato Ikeda, Eisaku Sato, Nobusuke Kishi, and Shigeru Yoshida (Table 3, panel A), which makes Koizumi look bad by comparison.[35]

Shrinkage: tenure and iid noise. Finally, a large leader effect requires a high shrinkage factor ψ due to a high signal-to-noise ratio, which is one of the most important empirical determinants of the LS leader effect (Appendix Figure 2C). This happens with some top leaders due to long tenures averaging out noise: Park Chung Hee (19), Chiang Kai-shek (23) and Lee Kuan Yew (30). It is also important that these leaders had long periods in our sample when they were not in office, so we could distinguish leader effects from country effects.

Country-specific iid noise is another major factor determining the leader effect and its statistical significance. Low noise tends both to raise the LS leader estimate and lower its error, as discussed in Section 2. Other top leader estimates have shorter tenures than those above but a lower country-specific standard deviation of the error term, such as Emilio Medici in Brazil (in office 1970-74 during the "Brazilian Miracle").[36]

A low standard deviation of the error term (σ_ε) in established democracies also allows the detection of leader effects for little-remembered high- or low-growth leaders. Japan accounts for four of the highest-growth leaders (Ikeda, Sato, Kishi, and Yoshida) and one of the lowest-growth leaders (Koizumi). Similarly, low iid noise means that democratic leaders from Austria,

[35] Koizumi also has the misfortune of being in a region with very high growth rates.

[36] The credit for the "Brazilian Miracle" is often given to Finance Minister Delfim Netto rather than to Medici. This highlights how favorable our assumptions are to the leader growth hypothesis.

	Name	Country	LS Leader Estimate	RMS Error	Raw Growth PC Average	Shrinkage (ψ)	Sig 99%	Tenure	Dem	1st Year	Rank (Incl. Insignificant)
Table 3: Leaders with a Statistically Significant Growth Contribution (tenure≥3yrs, with complete growth data*)											
Panel A: leaders with a significant POSITIVE growth contribution											
1	Khama	BWA	3.25%	1.01%	9.56%	0.43	1	15	0	1966	#1
2	Ikeda	JPN	2.84%	0.73%	8.65%	0.50	1	4	1	1961	#2
3	Rodriguez Lara	ECU	2.52%	1.05%	9.14%	0.39	0	4	0	1972	#3
4	Kishi	JPN	2.48%	0.73%	7.81%	0.50	1	4	1	1957	#4
5	Sato	JPN	2.48%	0.61%	7.70%	0.65	1	8	1	1965	#5
6	Razak	MYS	2.44%	1.01%	7.47%	0.43	0	5	0	1971	#6
7	Papadopoulos	GRC	2.43%	0.83%	8.21%	0.62	1	6	0	1968	#7
8	Medici	BRA	2.41%	0.94%	8.71%	0.51	0	4	0	1970	#8
9	Chun Doo Hwan	KOR	2.35%	0.85%	8.09%	0.60	1	7	0	1981	#9
10	Masire	BWA	2.21%	0.99%	5.87%	0.46	0	17	0	1981	#10
11	Lee Kuan Yew	SGP	2.11%	0.73%	6.26%	0.71	1	30	0	1961	#11
12	Raab	AUT	2.00%	0.51%	5.89%	0.75	1	8	1	1953	#12
13	Ahern	IRL	1.98%	0.55%	5.32%	0.72	1	11	1	1997	#13
14	Chissano	MOZ	1.96%	0.92%	4.26%	0.53	0	18	0	1987	#14
15	Chiang Ching-Kuo	TWN	1.94%	0.73%	6.82%	0.71	1	10	0	1978	#15
16	Rumor	ITA	1.82%	0.63%	6.85%	0.63	1	4	1	1969	#16
17	Hee Park	KOR	1.80%	0.66%	6.42%	0.76	1	19	0	1961	#17
18	Zhivkov	BGR	1.75%	0.66%	5.08%	0.76	1	19	0	1971	#19
19	Karamanlis	GRC	1.72%	0.68%	4.87%	0.75	0	13	0	1956	#20
20	Anerood Jugnauth	MUS	1.63%	0.80%	4.22%	0.41	0	17	1	1982	#24
21	Santer	LUX	1.60%	0.69%	4.72%	0.56	0	10	1	1985	#26
22	Yoshida, Shigeru	JPN	1.55%	0.78%	6.29%	0.43	0	3	1	1952	#29
23	Vargas	COL	1.49%	0.76%	2.22%	0.68	0	4	0	1987	#30
24	Bratteli	NOR	1.33%	0.66%	4.15%	0.60	0	4	1	1971	#38
25	Chiang Kai-shek	TWN	1.31%	0.61%	5.51%	0.79	0	23	0	1952	#39
26	Fanfani	ITA	1.31%	0.59%	4.96%	0.67	0	5	1	1958	#40
Panel B: leaders with a significant NEGATIVE growth contribution (rank from last)											
-1	Cedras	HTI	-3.53%	1.10%	-9.85%	0.333	-1	3	0	1992	-#1
-2	Lon Nol	KHM	-3.02%	1.18%	-11.12%	0.232	0	4	0	1971	-#2
-3	Amin	UGA	-2.74%	0.88%	-2.97%	0.576	-1	8	0	1971	-#3
-4	Gouled Aptidon	DJI	-2.67%	0.93%	-4.28%	0.520	-1	22	0	1977	-#4
-5	Mello	BRA	-2.34%	1.00%	-2.97%	0.449	0	3	0	1990	-#5
-6	Khalifah Ath-Thani	QAT	-2.13%	0.94%	-3.19%	0.510	0	23	0	1972	-#6
-7	Bandaranaike, S.W.F	LKA	-2.12%	1.04%	-2.85%	0.405	0	4	0	1956	-#7
-8	Ferenc Gyurcsany	HUN	-2.07%	0.85%	2.48%	0.604	0	4	0	2005	-#8
-9	Junichiro Koizumi	JPN	-2.06%	0.66%	1.08%	0.593	-1	6	1	2001	-#9
-10	Mobutu	COD	-1.98%	0.84%	-3.20%	0.612	0	31	0	1966	-#10
-11	Arbenz Guzman	GTM	-1.86%	0.79%	-0.99%	0.657	0	4	0	1951	-#11
-12	Mahendra	NPL	-1.63%	0.78%	0.41%	0.664	0	11	0	1961	-#18
-13	A. Papandreou	GRC	-1.52%	0.73%	0.68%	0.709	0	10	0	1982	-#21
-14	Wilson	GBR	-1.46%	0.56%	0.78%	0.709	-1	7	1	1965	-#23
-15	Marcos	PHL	-1.36%	0.69%	0.85%	0.739	0	20	0	1966	-#32
-16	Nixon	USA	-1.34%	0.58%	1.83%	0.686	0	6	1	1969	-#34
-17	Eisenhower	USA	-1.34%	0.53%	1.23%	0.734	0	8	1	1953	-#36
-18	Chirac	FRA	-0.81%	0.34%	1.72%	0.891	0	12	1	1995	-#108

Notes: list of leaders (tenure >=3yrs, with complete growth data) who have a significant LS leader effect at the 5% level. That is where (LS leader estimate)-1.96*(RMS Err)>0 or (LS leader estimate)+1.96*(RMS Err)<0. Sig at 99% is analogous. *PWT9 Growth data runs 1951-2014 so complete grwoth data excludes leaders with part of their tenure before 1951 or after 2014.

Luxembourg, Ireland, and Italy turn up on the positive growth leader list, while Eisenhower, Nixon and Jacques Chirac surface on the negative one.

Other factors. As noted above, our model is intentionally favorable to leaders, granting the presumption that they matter for growth in general. Other growth factors beyond leaders' influence that are not controlled for in our framework—for example, because they have heterogenous impacts or are difficult to measure—will affect the raw leader growth average. But their impact will also be shrunk toward zero by the shrinkage factor. This again highlights the difficulty of detecting which leaders matter based on the raw leader growth average (and on growth data in general).

Overall best and worst leaders for growth. As noted earlier, our overall best leader for growth is Seretse Khama, who led Botswana from independence in 1966 until his death in 1980. Acemoglu et al. (2003) report that Botswana had the world's highest average per capita growth rate in the 35 years before 2000, partly due to institutions and partly to "a number of important and farsighted decisions by the post-independence leaders." Khama benefits from a very high average per capita growth rate of around 10%, a long tenure of 15 years (which increases ψ), mediocre contemporaneous growth in other African countries, and slower (though still impressive) growth under other post-independence Botswanan leaders. The distribution of Khama's leader effect, conditional on observing his leader growth average $\bar{g}_{\mathrm{i}}$ and average growth under other Botswanan leaders $\bar{g}_{-\mathrm{ic}}$, is shown in Figure 1. The leader with the second highest significant growth contribution is Ikeda, who led Japan during the first half of the 1960s. He targeted doubling per capita income in a decade and introduced tax, trade, and spending policies to achieve that goal.[37]

[37] Takafusa Nakamura, an economic historian, described Ikeda as "the single most important figure in Japan's rapid growth. He should long be remembered as the man who pulled together a national consensus for economic growth."

Again, the leader with the most negative significant growth contribution (with at least 3 years of tenure) is Cédras, the general who led the September 1991 coup against Aristide in Haiti. The coup prompted a series of international trade sanctions that crippled Haiti's economy in the next three years when Cédras was the de facto leader, resulting in growth of -7%, -8%, and -15% (respectively) according to PWT9—the country's worst years on record. So it is still true that an extremely bad growth average can contribute to a significantly negative leader estimate. However, extreme growth is not sufficient. An even worse growth outcome (-16% on average for 3 years) occurred in Chad under Goukouni Oueddei, for whom we fail to reject a leader effect of zero. The difference is that the standard deviation of the iid noise is much higher in Chad than in Haiti. Cambodia leader Lon Nol (in power in the early 1970s) has the second most negative significant growth effect.[38]

Famous leaders. The good news for the "benevolent autocrat" hypothesis is that we do confirm significant positive estimated leader effects for celebrated leaders such as Khama, Lee, Park, and Chiang Kai-Shek. The bad news is that other celebrated leaders fail to show a significant leader effect, such as Deng, while uncelebrated leaders appear instead in the top ranks. The situation is similar in the table of negative-growth leaders, where famous disasters like Amin and Mobutu Sese Seko mix with little-known leaders.

Many famous leaders miss out because high iid noise results in small shrinkage factors (ψ) and large errors around leader estimates. For example, Deng's estimate also suffers from very high noise in China (a standard deviation of the error term of more than 6 percentage points) and Meles

[38] Lon Nol's regime was embroiled in a civil war against the Khmer Rouge, resulting in very negative growth rates. While his policies likely influenced the war, so did external factors beyond his control.

Zenawi (Ethiopia) has only a modest leader estimate and a high estimated error because of high iid noise.[39]

Many famous high-growth or low-growth leaders have less good (or bad) performances compared with other leaders of their countries. For example, Deng's insignificant leader estimate (rank #186 with tenure ≥ 3yrs) suffers from the high growth of most other Chinese leaders.[40]

We conjecture that the failure to confirm some famous successes and disasters is mostly because the attention paid to benevolent and malevolent leaders in autocratic countries is usually based on raw growth averages of leaders.[41] As we have observed, estimating the leader effect requires major adjustments to that raw growth average. Although raw growth averages can sometimes contribute to a strong leader effect (as for Khama or Cédras), they are not generally a good predictor of our set of significant high-growth or low-growth leaders in Table 3. Of the 44 significant high-growth or low-growth leaders, only 15 are in the best 25/worst 25 according to raw growth. Twelve of our leaders with a significant growth contribution are not even in the best 100/worst 100 by raw growth. This is yet more confirmation in the actual data of what we demonstrated with Monte Carlo simulations in Table 1: judging leader quality by raw growth averages leads to large errors relative to the optimal LS estimates of leader effects.

Alternative methodologies. This paper proposes a new methodology to estimate the growth contribution of individual national leaders. While the core approach should be uncontroversial

[39] For low-growth leaders, estimates for Khomeini (the Islamic Republic of Iran) and Somoza (Nicaragua) suffer from high noise as well as from the poor growth performance under other leaders.

[40] Setting $\gamma = 0$ increases the ranking of Chinese leaders, but they are still insignificant due to high noise.

[41] However, we do acknowledge that some of the differences reflect the inherent narrowness of the value-added approach. For example, our leader value-added approach would not have given enough credit to leaders like Nelson Mandela or Deng Xiaoping, who may have permanently changed the nations they led. Combined with our other results, this provides another reason against overreliance on contemporary growth data when evaluating leaders.

because it is optimal in a least-squares sense, the literature provides little guidance on modeling choices such as how long it takes for leaders to affect growth or which regional comparison groups should be used. Both require judgment calls. Naturally, the set of significant leaders will change depending on these assumptions, particularly for leaders with borderline significance. This reinforces our finding that it is very difficult to rigorously identify leaders who are good or bad for growth, and most leader estimates have wide standard errors. However, the fraction of statistically significant leaders—and the fraction of those who are democratic—is quite robust. To illustrate this, we now discuss the effects of two alternative assumptions.

The first alternative assumption we consider regards timing: that leaders do not affect growth contemporaneously as above but rather affect growth with a one-year lag.[42] In this case, the fraction of significant leaders falls slightly, from 6.8% to 5.1%, and the fraction of significant leaders who are democratic increases slightly, from 36% to 42%. Long-tenured leaders are less affected by the timing adjustment, whereas significant leaders with short tenures often lose significance (Online Appendix 5). The reduced fraction of significant leaders is consistent with a connection between leaders and growth that is stronger contemporaneously than lagged (estimates of σ_μ are also larger when estimated contemporaneously).

The second group of alternative assumptions regards the counterfactual growth rate against which leaders are compared: we require high-contribution leaders to achieve higher growth than

[42] One might argue instead that leaders have more persistent effects on growth, for example, by implementing reforms that change the long-run growth trajectory. While those claims are mostly untestable statistically, at a minimum we can ensure that leaders' own policies do not affect the estimates of country effects they are compared against. To do this, we simply calculate $\bar{g}_{-i}$ for each leader using only previous growth rates. If we do that, our results are mostly unchanged: 6.5% of leaders are significant, 38% of significant leaders are democratic, and the overall highest- and lowest-growth leaders are the same as in Table 3 (not reported).

other leaders in the same country, and higher growth than other leaders in the same region at the same time. If we instead compared leaders to an absolute standard of ≈2% global trend growth, the fraction of significant leaders is very similar (7.7%), as is the fraction of significant leaders who are democratic (42%).[43] Moreover, the overall highest- and lowest-growth leaders are the same as in Table 3. However, there is an overall shift towards leaders with positive significant contributions, driven by a greater fraction of leaders of European countries during the postwar boom and leaders of Asian countries, as these are no longer held against a higher regional growth standard.[44]

5. Conclusions

In this paper, we start with a growth model where leaders matter for growth in general. We combine growth data and an optimal signal extraction methodology to estimate the size and significance of the growth contribution of each individual national leader since 1950. Many of these significant leaders are little known or long forgotten. We show that our optimal methodology strongly dominates the naïve method of attributing individual leader growth averages to leaders.

We find that only a small fraction of leaders (around 7%) have significant positive or negative growth contributions: knowing that leaders matter for growth *in general* is very different from knowing *which* leaders matter for growth. Our findings suggest much more caution is needed in policy debates that attribute growth to leaders. We do find significant leader effects for some famous "benevolent autocrats" and "bad emperors." However, leaders in non-autocratic countries

[43] Specifically, here we do not adjust for region X year observables in the first stage and we set $\gamma = 0$ for all leaders.

[44] Using WDI growth data produces a similar fraction of significant leaders (6.6%) and democrats (31%). PWT 7.1 produces smaller estimates of σ_{μ}^{AUT}, which reduces the fraction of significant leaders to 3.7% and increases the fraction of democrats. The set of significant leaders varies across data sets: see Online Appendix 3 for a discussion.

are not overrepresented in the set of statistically significant leaders, even though they matter more for growth in general, because autocratic countries also have noisier growth processes that make it difficult to isolate true leader effects. We also show that a leader's growth rate in office is virtually useless by itself to identify good- and bad-for-growth leaders.[45]

References:

Aruoba, S.B., A. Drazen, and R. Vlaicu. 2019. "A Structural Model of Electoral Accountability." *International Economic Review* (forthcoming).

Ashworth, S., E. Bueno de Mesquita, and A. Friedenberg. 2018. "Learning about Voter Rationality." *American Journal of Political Science,* 62 (1): 37-54.

Acemoglu, D., S. Johnson, and J. Robinson. 2003. "An African Success Story: Botswana" in *In Search of Prosperity: Analytic Narratives on Economic Growth*, edited by Dani Rodrik, 80-119. Princeton: Princeton University Press.

Besley, T., J. Montalvo, and M. Reynal-Querol. 2011. "Do Educated Leaders Matter?" *Economic Journal*, 121, F205–F227.

Blinder, Alan S. and Mark W. Watson. 2016. "Presidents and the US Economy: An Econometric Exploration." *American Economic Review*, 106 (4): 1015-45.

Berry, C. and A. Fowler. 2018. "Leadership or Luck? Randomization Inference for Leader Effects (RIFLE)." *Mimeo*, University of Chicago.

Brown, C. 2019. "Economic Leadership and Growth." Working Paper

Chetty, R, J. Friedman, and J. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104 (9): 2593-632.

de Luca, G., A. Litina, and P. Sekeris. 2015. "Growth-friendly Dictatorships." *Journal of Comparative Economics* 43: 98–111.

[45] Interesting areas for future research include investigating whether some leaders permanently shift growth, the extent to which leader effects can be extracted in real time (before they leave office), and leaders' broader contributions in other realms (such monetary policy, fiscal policy, inflation, conflict or democratization).

Easterly, W., M. Kremer, L. Pritchett, and L. Summers. 1993. "Good policy or good luck? Country Growth Performance and Temporary Shocks." *Journal of Monetary Economics*, 32 (3): 459-483.

Feenstra, R., R. Inklaar, and M. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review*, 105 (10): 3150-3182.

Glaeser, E., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9 (3): 271-303.

Goemans, H. E., K. S. Gleditsch, and G. Chiozza. 2009. "Introducing Archigos: A Data Set of Political Leaders." *Journal of Peace Research*, 46 (2), (March): 269-183.

Heston, A., R. Summers, and B. Aten. 2012. "Penn World Table Version 7.1." Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Jones, B. and B. Olken. 2005. "Do Leaders Matter? National Leadership and Growth Since World War II." *The Quarterly Journal of Economics*, 120 (3): 835-864.

Kane, T., and D. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER WP 14607.

Marshall, M., T. Gurr, and K. Jaggers . 2016. "Polity IV Project, Political Regime Characteristics and Transitions, 1800-2015 Dataset Users' Manual." Centre for Systemic Peace.

Martinez, L. 2018. "How Much Should We Trust the Dictator's GDP Estimates?" *Mimeo*, University of Chicago.

Olson, M. 1993. "Dictatorship, democracy, and development." *American Political Science Review,* 87 (3): 567–576.

Rodrik, D. 2000. "Institutions for High-Quality Growth: What They Are and How to Acquire Them." NBER WP 7540.

Rothstein, J. 2010. "Teacher Quality in Educational Production: Tracking, Decay and Student Achievement." *The Quarterly Journal of Economics,* 125 (1): 175-214

Wooldridge, J. 2001. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 1st Ed.

Wolfers, J. 2007. "Are Voters Rational? Evidence from Gubernatorial Elections." Working paper [link].

Yao, Y. and M. Zhang. 2015. "Subnational Leaders and Economic Growth: Evidence from Chinese Cities." *Journal of Economic Growth,* 20 (4): 405–436.