Shrinking dictators: how much economic growth can we attribute to national leaders?

William Easterly and Steven Pennings[1]

May 2018

National leaders – especially autocratic ones - are often given credit for high average rates of economic growth while they are in office (and draw criticism for poor growth rates). Drawing on the literature assessing the performance of schoolteachers and a simple variance components model, we develop a new methodology to produce optimal (least squares) estimates of each leader's contribution to economic growth (correcting for noise, country effects and regional variation), and we calculate the precision of those estimates. We find that (i) only a small fraction of leaders have a statistically significant positive or negative growth contribution: for the vast majority, we can't separate their contribution from zero; (ii) the average growth rate during a leader's tenure is a [forecast] biased and inaccurate measure of their true growth contribution and for the majority disregarding their growth performance and assuming a zero effect is more accurate; (iii) democratic leaders (not autocratic leaders) are overrepresented in the set of significantly good or bad leaders, largely due to less noisy growth outcomes and (iv) while we do sometimes find sizable growth contributions of celebrated "benevolent autocrats", we also find that they are regularly outranked by other less celebrated leaders. Combined, the results cast doubt on our ability to identify good and bad leaders – especially autocratic ones – based on their growth performance.

[1] Easterly: NYU and NBER, Pennings: World Bank. Email: william.easterly@nyu.edu and steven.pennings@nyu.edu or spennings@worldbank.org The views expressed here are the authors', and do not 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, Ross Levine, and Laura Trucco, and participants in the NYU Development Seminar, NYU Development Research Institute Annual Conference, NEUDC, Mid-west political science conference, World Bank DEC internal seminar, Berkeley Haas Political Economy Seminar and Oxford CSAE Conference. Jorge Guzman provided assistance with data. Web: http://williameasterly.org/ and https://sites.google.com/site/stevenpennings/

Section 1: Introduction

Both popular and academic discussions often give credit for high growth to the leaders presiding over that growth. In democratic countries this is part of the day-to-day partisan political debate. Leaders get even more credit or blame in autocratic countries, because a lack of constraints on executive power arguably make economic performance more sensitive to the intentions and abilities of national leaders. What can more formal methods say about attributing growth outcomes to leaders? This paper shows that even if we do give leaders credit (or blame) for growth, we can do much better than simply assigning them the average growth rate during their time of office.

This literature addresses a long-standing debate in development about "benevolent autocrats," by both academic and non-academic commentators. This debate has often made claims about particular leaders. For example, the *New York Times* obituary for Deng Xiaoping asserted "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."[2] More recently, *The Economist* magazine in March 2016 describes Rwanda as development's "shining star" with "average growth of 7.5% over the past 10 years," suggesting "much of its success is due to effective government" under Paul Kagame. It quotes diplomats as worrying that "without Mr Kagame's firm hand … the miracle wrought in Rwanda could quickly be reversed."[3] Particular high growth autocrats such as Kagame often emerge as an "aid darling," as foreign aid donors seem willing to overlook the leader's political repression because of the high growth he produces (Curtis 2015).

The discussion usually acknowledges that autocrats can be disasters also, so one popular position is that the variance of leaders as miracles or disasters is higher under autocracy than democracy. Glaeser et al. (2004) stress under autocracy the importance of "choices of their – typically unconstrained – leaders," noting the large variation possible across dictators: "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." De Luca et al (2015) analyze how some dictators will be "growth-friendly dictators" 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). Other dictators that lack an encompassing interest in the national economy will be more likely to destroy the economy if it maximizes their own gains from doing so.

Rodrik (2000) summarized the stylized consensus on the greater variability of leader growth effects under autocracy compared to democracy: "living under an authoritarian regime is a riskier gamble than living under a democracy."[4]

The academic literature on national leaders and economic growth goes back to Jones and Olken (2005) asking how much of the *general* variation in growth can be attributed to changes in national leadership. Utilizing unexpected leader deaths in office — such as by illness or accident — Jones and Olken found that underlying growth changes significantly across transitions for autocrats (in either direction), but not for democrats. Many papers follow Jones and Olken in testing for general leader effects, ranging from how they vary with the leader's education (Besley et al 2011); leader's education in economics (Brown 2017), across leader transitions like coups (Meyersson 2016), in

[2] http://www.nytimes.com/learning/general/onthisday/bday/0822.html

[3] http://www.economist.com/news/middle-east-and-africa/21694551-should-paul-kagame-be-backed-providing-stability-and-prosperity-or-condemned

[4] Rodrik (2000) finds that the within-country variability of growth is higher under autocracies, which is consistent with strong autocratic leader effects (though he does not test leader effects themselves).

different contexts, like city mayors (Yao and Zhang 2015), across different political parties (Blinder and Watson 2016) or using alternative methodologies (Berry and Fowler 2018).

In contrast with estimating general variation in growth due to leaders, in this paper we ask a different question: how much growth can we quantitatively attribute to each *particular* leader?

The focus on the growth effect of particular leaders is important because discussions of national leaders are rarely in the abstract and usually pass judgement on particular leaders. For example, Jones and Olken (2005) motivate their paper in part as a test of the "Great Man" view of history, where "history is largely determined by the idiosyncratic, causative influences of certain individuals, and perhaps a very small number" (p838). Aid policymakers must decide whether to support a specific leader, weighing their pro-growth benevolence against their negative tendencies like human rights violations.

This paper provides a method for measuring leader effects in an optimal way that minimizes the squared errors between the true leader quality and our estimate. It discusses how to quantify the precision of that estimate. We compare our method with other simple rules-of-thumb for estimating leader effects like assuming leaders have no effect on growth, or giving leaders credit for all growth during their tenure.[5] We also provide a list of statistically significant best and worst leaders for growth implied by the optimal methodology.

The problem of assessing the contribution of national leaders to growth is similar to the problem of assessing the "value added" of a school teacher to standardized test scores – about which there is a voluminous literature (e.g. Kane and Staiger 2008, Chetty et al 2014 among many others). To our knowledge, we are the first to make this connection.[6]

The key insight is that the leader growth average (like students' test scores under a particular teacher) is a noisy indicator of true performance. A high leader growth average is often due to good luck in addition to any good policy on the part of the leader. Hence our best estimate of true leader quality (or teacher quality) shrinks the growth average towards zero to adjust for the contribution of luck to the leader growth average. In the leader growth context we also need to adjust for country-specific and observable regional determinants of growth that are beyond the leader's control. Finally, we need to quantify precision via confidence intervals on the contribution of each leader (we will actually use the Bayesian analogue, credible intervals (CI)). Our method offers a useful new tool to optimally extract information about the leader's quality from growth data and evaluate its accuracy.

We will show all three of these corrections – noise, country effects, and observables – have a major effect on estimates of the size of individual leader effects on growth, the width of CIs, and the set of significant best and worst leaders. This leads to four main results.

First, we find that less than 25 leaders have a "statistically significant" positive growth effect – and less than 25 leaders have a "statistically significant" negative growth effect – out of around 750 leaders with tenures of at least 3 years in our data. As such, for the vast majority of leaders we can't distinguish their growth contribution from zero using available growth data, despite using an optimal signal extraction methodology, assuming a growth process innately favorable to leaders, and estimating variation in underlying leader quality sometimes larger than in Jones and Olken (2005). This means it's very difficult to identify either a "benevolent autocrat" or a "bad emperor", such

[5] The latter is equivalent to regressing of growth on leader dummies, and allocating the leader the estimated coefficient.

[6] We are grateful to Hunt Allcott for making this suggestion.

that policymakers and commentators should be very careful opining on who are the best and worst leaders for growth.

Second, our results show that the average growth rate during a leader's tenure is very misleading as to the leader's true effect on growth -- even under our assumptions that are favorable to that hypothesis. Using the leader growth average to measure the leader effect has a strong forecast-bias, has a root mean squared error that is more than double that of the optimal methodology, and fails to correctly identify good and bad leaders. The optimal method involves "shrinking" down the average growth rate by a factor of four. The leader growth average is so bad as a measure of leader quality that for the majority of leaders it is more accurate to simply ignore it and assume that the leader had no effect on growth at all (in either direction).

Third, while we do find evidence of higher underlying variation in leader quality in autocracies and transition countries than in established democracies (as do JO), this does not mean the best and worst observed leader list will be dominated by autocrats, as one might have expected under the "risky bet" hypothesis. In fact, because established democracies have much less noisy growth processes, leaders from those countries are *overrepresented* in the set of best of best and worst leaders – in part because of less shrinkage, and in part because of greater precision in the estimating their contributions.

Finally, while our estimates confirm some famous good or bad leaders, they also omit many others and mostly feature surprising, unknown or forgotten leaders. As such, our list of significantly best and worst leaders is quite different from that in the previous academic or policy discussions - and from our own priors.

The rest of the paper is organized as follows: Section 2 outlines our least squares methodology and explains how it works, with Section 3 discussing the data. Section 4 presents Monte Carlo evidence that our method outperforms alternatives. Section 5 estimates underlying variation in leader quality. Section 6 describes the variation in least squares leader effects, with Section 7 presenting the best and worst leaders. Section 8 concludes. The online appendix contains proofs (Appendix 1), additional robustness tests using different growth datasets (Appendix 2), and extra figures and tables (Appendix 3).

Section 2: Model and Methodology

Section 2.1: Estimates of the optimal (least squares) leader effects

In the academic literature and in policy discussions, leaders are often attributed the average growth during their tenure. This is equivalent to estimating leader effects by regressing growth on dummy variables for each leader, i.e. doing fixed effects in a panel regression. There are still three problems with this approach. First, and most importantly, the random *idiosyncratic* noise component of growth is very large (Easterly et al 1993 and many papers since) and tends to swamp leader effects even over the medium term. A unusually high growth average under a leader would be completely attributed to the good (or bad) policies of a leader, when it is likely to also include a positive noise realization . Second, some *countries* have higher or lower trend growth rates due to other factors that are not related to individual leaders – such as institutions, culture or geography.[7] Finally, there are *other supranational forces* that can affect growth, such regional economic trends or commodity price shocks – Latin

[7] Institutions that consistently select good leaders, or constrain bad ones, would come through as part of the country effect. Initial conditions, like the stock of human capital, might also contribute to the country effect (Easterly and Levine 2016).

American countries generally grew more slowly during the "lost decade" of the 1980s than in the 1950s or 1960s, and sub-Saharan Africa has grown more slowly than Asia on average.

This simple statistical model of growth is summarized in Equation (1). Annual per capita GDP growth g^*_{ict} under leader i in country c, during year t can be decomposed into a leader contribution (μ_i), a vector of observables including regional time dummies (for regional averages and business cycles) in X, a country-specific component (μ_c) and idiosyncratic error (ε_{ict}) component for a panel of leaders:

(1) $g^*_{ict} = X\delta + \mu_i + \mu_c + \varepsilon_{ict}$

$X\delta$ are observable, which means that we can control for them via regression, leaving the growth residual g_{ict} (Equation 2). g_{ict} then depends on three unobserved random variables μ_i, μ_c, ε_{ict}, from which the country draws $\mu_c \sim (0,\sigma_c^2)$, each leader draws a quality $\mu_i \sim (0,\sigma_\mu^2)$ and for each period $\varepsilon_{ict} \sim (0,\sigma_{c\varepsilon}^2)$, with μ_i, μ_c and ε_{ict} being independent and serially uncorrelated.

As in Jones and Olken (2005), a crucial part of this model is that the leader quality for a particular leader is a random draw from a distribution (a random effect), and not a non-random universal parameter to be estimated (a fixed effect). This changes the econometrics of estimating a leader effect considerably, and means that we can no longer rely on classical asymptotics like the central limit theorem. Because our estimation target is a realization of a random variable rather than a parameter, classical frequentist concepts like bias and confidence intervals don't apply, though we can replace them with analogous Bayesian concepts of forecast bias and credible intervals.

Note that country effects represent differences from regional averages because these were already removed as part of $X\delta$ controls. We assume ε_{ict} is independent across years and countries, but we allow it to be heteroskedastic by country. This turn out to be crucially important for our results, as many of the countries with the most extreme leader growth averages tend to have very noisy growth processes. Each leader is in power for T_i years and N_c is the total number of years in the sample for country c.[8]

(2) $g_{ict} \equiv g^*_{ict} - X\beta = \mu_i + \mu_c + \varepsilon_{ict}$

Note that we are intentionally modeling growth to be as favorable as possible to the practice of attributing growth to leaders. We give leaders full credit for *all* growth during their tenure except for that due to observable international factors (regional growth differences, regional business cycles, etc), country effects and iid shocks. Other time-varying but persistent factors that affect growth will be attributed to leaders and bias upwards the absolute size of leader effects. Hence our exercise provides an upward bound on the (absolute) size of contemporaneous leader effects.[9]

[8] Equation (2) is very similar to that in Jones and Olken (2005, p840) with a few differences in notation. JO denote the SD of leader quality as σ_l (which they normalize to 1), the effect of one unit of leader quality on growth is θ, a parameter where they test if $\theta \neq 0$. This means the SD of the leader effect on growth is $\theta\sigma_l = \theta$, is equivalent to σ_μ here (hence, JO effectively test if $\sigma_\mu = 0$). Our country effect μ_c is ν_i in JO (which is non-stochastic), and they allow for the leader quality to be serially correlated.

[9] Note, however, that we don't include any effect of leaders (good or bad) after they left office. For example, when George Washington retired from office after two terms he created a powerful precedent that likely saved the United States economically costly leadership struggles in the future (and formed the basis of the 22nd amendment). But this would not be included in our estimate of Washington's leader effect (it would be close to impossible to estimate its size anyway).

We define *the leader residual growth average*, $\bar{g}_{ic}$, as the average of the *residual* growth rate of leader *i* in country *c* with tenure T_i , after we remove observables such as regional business cycle (which includes the mean growth rate) as given by Equation (3):

$$(3) \quad \bar{g}_{ic} = \frac{1}{T_i}\sum_{t=1}^{T_i} g_{ict}$$

It will also be useful to record the average *residual* growth rates for all *other* leaders than *i* in the same country c (which we denote –*i*) after removing observables, which we use as a measure of country effects.

$$(4) \quad \bar{g}_{-ic} = \frac{1}{N-T_i}\sum_{t=T_i+1}^{N} g_{-ict}$$

We calculate the growth residual used in Equation (3) and (4) by running regressions on observables such as time by region fixed effects to capture the regional averages and business cycles, which also removes the mean worldwide growth rate (which is around 2%) from the leader growth averages.[10] [11]

Formal definition of problem

We want to have the "best" estimate of the size of the (unobservable) leader effect μ_i based on *observable* data: the average growth rate residual during that leader's tenure $\bar{g}_{ic}$, and the average growth rate residual under *other* leaders in the same country $\bar{g}_{-ic}$. "Best" here is the minimum squared error, as commonly used for evaluating forecasts (in the Monte Carlo simulations we report the Root Mean Squared Error RMSE). We also restrict the model such that the leader estimate $\hat{\mu}_i$ is a linear function of the own leader growth average and leader growth average of *other* leaders $\hat{\mu}_i = \beta_1\bar{g}_{ic} + \beta_2\bar{g}_{-ic}$. This can be rearranged into a more intuitive form (without making any restrictions), as in Equation 5. $\bar{g}_{ic} - \gamma\bar{g}_{-ic}$ is the *adjusted* leader growth average which uses the economic performance under *other* leaders $\bar{g}_{-ic}$ as a proxy for the country effect. ψ is the *shrinkage factor* which adjusts downward the adjusted leader growth.[12] γ determines how useful the other leader growth average is for identifying country effects, and weights the $\bar{g}_{-it}$ accordingly.

$$(5) \quad \hat{\mu}_i = \hat{\psi}(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})$$

The problem is to choose ψ and γ to minimize the expected squared error:

$$(6) \quad \min_{\psi,\gamma} E\left[\mu_i - \psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})\right]^2$$

[10] In previous versions of this paper we also added a country-specific commodity export price index and a conflict dummy variable to the set of observables, which generated very similar results (not reported).

[11] Although our optimal methodology is general to allow it, we usually don't want to subtract country-specific means because we cannot separately identify a country effect from a string of leader effect – especially if there are few leaders in the country.

[12] $\psi = \beta_1$ and $\beta_2 = -\gamma\psi$

Definition D1 ***Forecast Unbiasedness*** Following Chetty et al (2014), we call leader effect estimator *forecast unbiased* if for a regression of the form $\mu_i = \lambda\hat{\mu}_i + e_i$ then $\hat{\lambda} = 1$.[13]

In other words, an estimator is forecast unbiased if a 1ppt increase in measured leader quality $\hat{\mu}_i$, on average is associated with a 1ppt increase in true (unobserved) leader quality μ_i.[14]

Note that *forecast bias* is a different concept from classical bias (that in repeated samples the estimator averages out to the true parameter) and one that is more applicable in our context. The classical definition makes little sense here because μ_i is a one draw of a random variable rather than a being a fixed parameter: in repeated samples one would also get a different μ_i, with both estimator, leader effects and noise averaging out to zero. If one applied a classical approach to our shrinkage estimator (with a fixed leader effect), one would find that it was biased towards zero. This is a *feature* of all shrinkage estimators: that a reduction in RMSE can be obtained by reducing variance of the estimator via shrinkage – even if this results in a small amount of classical bias.

Proposition 1 ***Least-squares leader estimate***
Consider the growth model given by Equation 2 and where the leader estimate is a linear function of own and other leader growth average residuals (Equation 5).

(A) Then the values $\hat{\psi}$ *and* $\hat{\gamma}$ *which minimize the expected squared leader error (Equation 6) are given in equations (7) and (8).*

$$(7)\quad \hat{\gamma}_i = \frac{E(\bar{g}_{ic}\bar{g}_{-ic})}{E(\bar{g}_{-ic}^2)} = \frac{\sigma_c^2}{\sigma_c^2 + \frac{\sigma_e^2}{N-T_i} + \frac{\sigma_\mu^2}{L_{-ict}}}$$

$$(8)\quad \hat{\psi} = \frac{E\left[\mu_i(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})\right]}{E((\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})^2)} = \frac{\sigma_\mu^2}{\sigma_c^2(1-\hat{\gamma}) + \sigma_\mu^2 + \frac{\sigma_e^2}{T_i}}$$

(B) The resulting estimator (with $\hat{\psi}$ *and* $\hat{\gamma}$ *defined in Eq 7 and 8) is forecast unbiased (as in definition D1)*

Proof: See Appendix.

The optimal $\hat{\gamma}$ increases with σ_c^2 -- the population variance of the country effect (which increase the strength of the signal in $\bar{g}_{-it}$) -- but decreases in σ_e^2 (the variance of the iid error term which varies by country). It also decreases with σ_μ^2 (the variance of the leader effect), which make the other leader growth average a nosier measure of the country effect. We use the notation T_i as the tenure of this leader, in case it is different from the average tenure of other leaders in the country (T_{-i}). L_{-ict} is the number of other leaders in the same country.

If σ_c^2 =0 then $\hat{\gamma} = 0$, and the *adjusted* leader growth average is just the leader growth residual average - there is no need to adjust for country effects when there aren't any (we call this the simple model below).

[13] We do not include a constant here because leader estimates and all of the estimators we consider are mean zero.

[14] If $\hat{\lambda} \neq 1$ (as we estimate for the leader growth average), then the larger the leader estimate, the larger the error.

In contrast, when $\sigma_e^2 = \sigma_\mu^2 = 0$ and $\sigma_c^2 > 0$, then the average growth rate under *other* leaders is a perfect signal of the size of the country effect μ_c and so $\hat{\gamma} = 1$. $\hat{\gamma}$ is also close to one if (i) there is a long sample for the country (a large N_c) which smooths out the iid noise and (ii) there are many other leaders in the same country (a larger L_{-ict}), so that the other leader effects even out. If $\hat{\gamma} \approx 1$ then we subtract the full *other* leader average residual from the leader growth average residual. In countries where growth was high under other leaders, the model will attribute most of this to the country effect and adjust the leader growth average downwards.

The estimate of the *shrinkage* factor is given by Equation (8), which is the weight applied to the adjusted leader growth residual average ($\bar{g}_{ic} - \gamma \bar{g}_{-ic}$). One can see that if $\sigma_c^2 = \sigma_e^2 = 0$, then $\hat{\psi} = 1$ and the best estimate of the leader effect μ_i is $\bar{g}_{ic}$ - the leader growth average residual is a perfect signal of the leader's effect on growth.

However, if $\sigma_e^2 > 0$ then the optimal shrinkage factor $\hat{\psi} < 1$ because the leader growth average includes noise due to the idiosyncratic shocks to growth ε_{ict}. $\hat{\psi}$ will be especially small due to the iid error if (i) leaders have a short tenure (small T_i), or if (ii) σ_ε^2 tends to be large. We will see that (ii) is the case more in autocratic than in democratic countries. For leaders with a long tenure, these random errors even out over time meaning the leader growth average is more informative of the true leader growth effect.

The shrinkage factor $\hat{\psi}$ will also be small in the case that the country effect σ_c^2 is large *and* we are not able to control for it very precisely by subtracting for the *other* leaders' average because $\hat{\gamma} << 1$. This might be the case if there the sample of other leaders is small (N_c - T_i is small).

Of course, an extremely high or low average growth under leader i ($\bar{g}_{ic}$) still matters, it can lead to a strong leader estimate even after being shrunk.

Section 2.2: Estimating the precision of leader estimates

In this subsection we calculate the credible intervals (the Bayesian analogue of confidence interval) around our least squares leader estimates. As leader estimates in Proposition 1 are least squares, by construction they have the highest precision (smallest root mean squared error, RMSE) of all estimators that are a linear function of $\bar{g}_{ic}$ and $\bar{g}_{-ic}$. Some of the RMSEs of some alternative rules for assigning leader effects (no leader effect $\hat{\mu}_i = 0\ \forall i$ and the naïve leader growth average $\hat{\mu}_i = \bar{g}_{ic}$), are calculated below in Section 2.3.

Measures of precision are somewhat different from a standard regression framework, so we derive the conditions explicitly in Proposition 2. The difference is because we are estimating a particular draw of a random variable based on a fixed sample size (of the leaders' tenure), rather than estimating a fixed non-stochastic parameter based on a potentially unlimited sample size (as in standard regression framework). This means we can't rely on asymptotics like the central limit theorem to provide a distribution for our estimator and the classical concept of a confidence interval doesn't apply (as they relate to estimates of a fixed parameter with repeated random sampling).

Instead we use an analogous concept of a Bayesian *credible interval*, which is based on the posterior probability distribution of the leader estimate μ_i, given the data. That is, a 95% *credible interval* is the range of values that the true leader effect μ_i can take with 95% probability (based on the posterior distribution of μ_i). It turns out that our least squares leader estimates are *also* the Bayesian maximum posterior estimates ($\max f(\mu_i|\bar{g}_i, \bar{g}_i)$ and Bayesian mean posterior ($E(\mu_i|\bar{g}_i, \bar{g}_i)$) estimates and that the RMSE is the standard deviation of the posterior distribution $f(\mu_i|\bar{g}_i, \bar{g}_i)$.[15] As the posterior is a normally distributed, this means that the width of the 95% credible intervals is c $1{,}96 \times RMSE$ around the leader estimate. These results stems from growth rates and errors being normally distributed (Assumption A1), which is a common assumption in the literature.[16] [17]

Assumption A1: The error components in Equation 2 are normally distributed with mean zero, that is $\mu_i \sim N(0, \sigma_\mu^2)$, $\mu_c \sim N(0, \sigma_c^2)$, $\varepsilon_{ict} \sim N(0, \sigma_{e,c}^2)$.
Assumption A2: the values of ψ_i and γ_i are known for leader *i*

As the sum of normal is also normal, a corollary of Assumption A1 is that the growth rate residual, the leader growth average $\bar{g}_{ic}$ and other leaders' growth average $\bar{g}_{-ic}$ are also normal.

Proposition 2 ***Precision ("Credible intervals") of Least Squares Leader Estimates***
Assume A1 and A2 then:
(A) The root mean squared error (RMSE) of the least squares leader estimate is given by equation 9:

$$(9)\quad RMSE(\hat{\mu}_i) \equiv \sqrt{E[\hat{\mu}_i - \mu_i]^2} = \sqrt{(\psi_i - 1)^2\sigma_\mu^2 + {\psi_i}^2[\sigma_e^2/T_i + (1-\gamma)\sigma_c^2]}$$

(B) The posterior is $\mu_i|\bar{g}_i, \bar{g}_i \sim \mathrm{N}(\hat{\mu}_i^{LS}, RMSE_i(\mu_i)^2)$ where $\hat{\mu}_i^{LS} = \psi_i(\bar{g}_i - \gamma\bar{g}_{-i})$ is the least squares leader estimate from Proposition 1.
(C) For any estimated least squares leader effect $\hat{\mu}_i^{LS}$:

$$\mathrm{P}(\hat{\mu}_i^{LS} - Z_{\alpha/2}^{\star} \times RMSE(\hat{\mu}_i) \leq \mu_i \leq \hat{\mu}_i^{LS} + Z_{\alpha/2}^{\star} \times RMSE(\hat{\mu}_i))=1\text{-}\alpha$$

where $Z_{\alpha/2}^{\star}$ is 1-$\alpha/2$ critical value from a normal dist. (ie if 1-α=95%, $Z_{\alpha/2}^{\star} \approx 1.96$; if 1-$\alpha$=99%, $Z_{\alpha/2}^{\star} \approx 2.576$).
Hence the 1-α% credible interval (CI) is $[\hat{\mu}_i^{LS} - Z_{\alpha/2}^{\star} \times RMSE(\hat{\mu}_i), \hat{\mu}_i^{LS} + Z_{\alpha/2}^{\star} \times RMSE(\hat{\mu}_i)]$.
Proof: See Appendix.

[15] From a Bayesian perspective (with a normal prior and normal leader growth averages), the estimated leader effect is a precision weighted average of the prior mean (of zero), and the data (the leader growth average), with the weights being the inverse of the variance of prior and noise of the leader growth average. When the growth average (data) is more noisy, a Bayesian approach puts more weight on the prior.

[16] A second complication is that the estimates of parameters ψ_i and γ_i are a nonlinear function of variances (rather than means) and so have a non-standard distribution. In order to calculate analytical errors (and distributions) we need to assume that ψ_i and γ_i are known (Assumption A2). In an extension, we relaxed this assumption and use a country-level block bootstrap to calculate higher moments of $\hat{\psi}_i$ and $\hat{\gamma}_i$, and simulation methods to estimate the empirical distribution of the error $\hat{\mu}_i - \mu_i$, taking the estimation error of $\hat{\psi}_i$ and $\hat{\gamma}_i$ into account (not reported). Despite being much more complicated than the analytical formulas in Proposition 2, the estimated errors, credible intervals and set of significantly good or bad leaders are almost unchanged. This suggests that parameter uncertainty is less important than underlying noise about year-to-year growth.

[17] For example Jones and Olken (2005) also assume that the iid errors and leader quality in Equation 2 were normally distributed. The distributional assumptions are not required to calculate RMSEs themselves, but are required to interpret $2 \times RMSE$ as the width of confidence intervals.

Definition D2 ***Statistical Significance*** We call a leader estimate $\hat{\mu}_i^{LS}$ *statistically significant*, if zero is outside the 95% (or 99%) credible intervals defined in Proposition 2C.[18]

Proposition 2A shows that the RMSE has two terms, reflecting the tradeoff between classical bias (first term) and the noisiness of the estimator (second term) when choosing ψ.[19] When $\psi < 1$, an error is introduced as 1- ψ of any leader effect μ_i is shrunk away, which is problematic when leader quality varies substantially (high σ_μ^2). This error is removed when $\psi \approx 1$, like for the leader growth average. However when $\psi \approx 1$, the second term is large, which creates errors whenever there is a lot of idiosyncratic noise during the leader's tenure (high σ_e^2 and short tenure T_i), or whenever country effects important (high σ_c^2). If we can identify country effects accurately using the other leader growth average, then $\gamma \approx 1$, and the final term $(1-\gamma)\sigma_c^2$ drops out.

Proposition 2C establishes the Bayesian credible interval that will be used to establish our set of "statistically significant" leaders (defined by definition D2). A Bayesian credible interval is based on the posterior distribution, so we need proposition 2B to characterize that distribution, to show that is centered on our least squares leader estimate $\hat{\mu}_i^{LS}$ (and to link to the RMSE).[20]

Section 2.3 Understanding least-squares leader estimates: comparison with simple rules and a simple model
Instead of calculating the least squares leader estimates as above, commentators might follow a simple rule-of-thumb that that attributes all or none of the leader growth average residual to the leader. We evaluate each of these "simple rules" below and compare to the least squares leader estimates.

<u>Naïve leader growth average</u>
The first simple rule is assigning each leader *all* of the average growth residual during their tenure, which we call the "naïve" rule.

(10) $$\hat{\mu}_{\mathrm{i}}^{naive} = \bar{g}_{it}$$

This is a restriction of the class of linear rules above where $\psi_i^{naive} = 1$ and $\gamma_i^{naive} = 0$. In this case the RMSE of the estimator is:

(11) $$RMSE\left(\hat{\mu}_i^{naive}\right) = \sqrt{\sigma_c^2 + \sigma_e^2/T_i}$$

(which simplifies to $\sigma_e\ /\sqrt{T_i}$ if there are no country effects). The shrinkage factor of $\psi_i^{naive} = 1$ and 2x RMSE bounds are plotted in Figures 1 and 2 respectively. The errors of the naïve rule are large whenever growth is very noisy (large σ_e) and tenures are short (small T_i). The rule also has a large error when country effects are large, and so the leader growth average reflects country-specific factors rather than leader-specific factors.

The naïve rule is also forecast biased (as in definition D1), with:

(12) $$\lambda_{naive} = \frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_c^2+\sigma_e^2/T_i} < 1$$

[18] Statistical significance is usually a classical concept, and so we define it specifically here.

[19] Let $E_\mu(.)$ and $var_\mu(.)$ be the classical expectation and variance operators treating μ_i as a fixed parameter. Let classical bias be $B_\mu = E_\mu(\hat{\mu}_i) - \mu_i$. Then $RMSE_\mu \equiv \sqrt{E_\mu((\hat{\mu}_i - \mu_i)^2)} = \sqrt{B_\mu^2 + Var_\mu(\hat{\mu}_i\ )}$, with shrinkage leading to large reduction in $Var_\mu(\hat{\mu}_i\ )$ at the cost a small increase in B_μ^2.

[20] As the posterior is normal, with mean $\hat{\mu}_i^{LS}$, the central and maximum-posterior credible intervals are the identical and are based around our least squares leader estimate.

Zero leader growth average

At the other end of the scale, one can adopt a rule of thumb that leader growth averages are too noisy, and so it is better to attribute *none* of the leader growth average as due to the leader – which we call the "zero" rule.

(13) $\hat{\mu}_{\mathrm{i}}^{zero} = 0$

which implies $\psi_i^{zero} = 0$ (the value of γ doesn't matter). In this case the RMSE of the estimator is:

(14) $RMSE(\hat{\mu}_i^{zero}) = \sigma_\mu$

The shrinkage factor and 2xRMSE bounds are plotted in Figures 1 and 2 respectively. The error of the zero rule comes from misses on good and bad leaders, and so generates large errors whenever leader quality varies substantially across leaders (large σ_μ), but doesn't depend on tenure, iid error or country effects. Forecast bias is not defined for the zero rule as $\hat{\mu}_i = 0$.

The RMSE, by construction, is larger for each of these simple rules than the least squares estimate above. But when is one simple rule-of-thumb preferred to the other? The RMSE of the zero rule is less than that of the naïve rule whenever $\sigma_\mu^2 < \sigma_c^2 + \frac{\sigma_e^2}{T_i}$, which can be rearranged $T_i < \frac{\sigma_e^2}{\sigma_\mu^2 - \sigma_c^2}$. If leader quality is just as variable as country quality, $\frac{\sigma_e^2}{\sigma_\mu^2 - \sigma_c^2} \rightarrow \infty$ and the zero rule is always preferred. With the variance components estimated below, $\frac{\sigma_e^2}{\sigma_\mu^2 - \sigma_c^2} \approx 10$, so for the 80% of leaders with tenure of a decade or less – it is more accurate to assume the leader of average quality (and ignore all growth data), than to attribute the entire growth average during the leaders' tenure to the leader.[21] Of course, our estimator still dominates both of these options regardless of this calculation.

Simplified model with no country effects

The most important determinants of the shrinkage factor – how noisy is the growth process, and leader tenure – can be illustrated with a simplified model with no country effect. Then Equation 8 can be applied directly to leader growth average residual, which simplifies to:

(15) $\psi_{i,simple} = \sigma_\mu^2 / (\sigma_\mu^2 + \sigma_e^2 / T_i)$

In the simple model, when the iid growth errors are large and the tenure is short, the shrinkage factor is close to zero, as one can see in Figure 1 (described below). Theoretically, as tenures become long the shrinkage factor approaches 1, though for realistic values of σ_u, σ_e and T_i the shrinkage factor is well below 1.

In the simple model, $\sigma_c = \gamma_i = 0$, and hence the error in Equation (9) simplifies to:

(16) $RMSE(\hat{\mu}_i)_{simple} = \sqrt{(\psi_i - 1)^2 \sigma_\mu^2 + {\psi_i}^2 \sigma_e^2 / T_i}$

Understanding the determinants of the leader effect

In Figures 1 and 2 we plot leader tenure vs the shrinkage factor or 2 x RMSE bounds (respectively), for the simple rules, full model least squares estimates, and a simplified model without any country effects ($\sigma_c^2 = 0$). In Panel A, we assume a high iid error variance (σ_e =5%) typical in autocratic countries, and in Panel B a low iid error variance (σ_e =2.65%) typical in democratic countries. All other parameters are identical to isolate the importance

[21] Specifically, the cutoff is 10 years for established democracies, 13 years for autocracies/transition countries and 17years for a pooled sample. These cutoffs are different from those in Figures 1 and 2 because we assume a common leader SD and country SD in those figures.

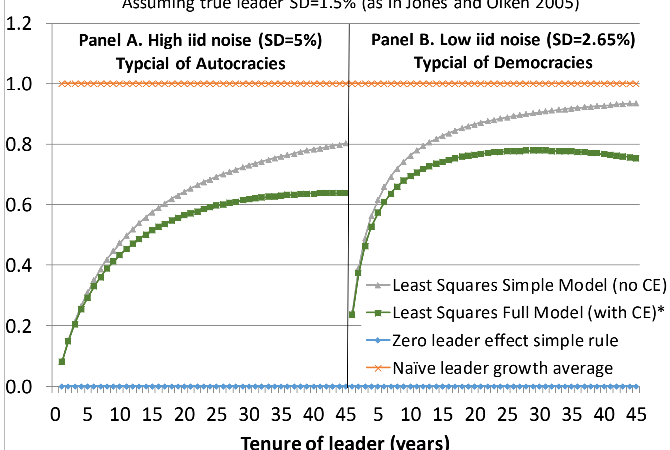
of the iid error variance in driving our results. Specifically, we assume that the *unobservable* standard deviation of the leader effect is $\sigma_\mu = 1.5\%$ across all leaders which is taken from Jones and Olken (2005 p837), and for the full model, that the country SD $\sigma_c = 1\%$ -- both similar to what is estimated below in the actual data. No leader in our dataset has tenure more than 45 years, with average tenure being around 5-6 years.

Figure 1 shows that that (i) least squares leaders shrinkage factor ψ is much smaller for high iid error countries than lower iid error countries and (ii) is increasing but inverse-u-shaped in the tenure of the leader. For example, for a leader of 5 years tenure (close to the average), the shrinkage factor is around 0.3 in high iid noise countries, which is around half of the 0.6 in low iid noise countries. Because autocracies tend to have much more noisy growth processes than democracies (it does not matter if this is causal), the average growth rate under autocrats is often a particularly *unreliable* estimate of the true leader contribution. For autocratic leaders with short tenures, the optimal shrinkage factor is very small, and so assuming a zero leader effect is not a bad approximation of the least-squares leader effect (and generates almost identical errors).

Figure 2 plots 2x RMSE bounds, which are the width of 95% credible intervals around least squares leader estimates and so reflect the absolute value of a LS leader effect that is marginally statistically significant. Most striking is how wide the 2x RMSE bounds are for the naïve simple growth average rule. In countries with noisy growth (like most autocracies), there is very little information in the leader growth average for tenures of less than 5 years, such that a leader that was averaging 7% per capita growth (5 ppts above the 2ppt cross-country mean), would have an contribution less than 2 RMSEs away from zero.

Figure 1: Shrinkage factor (ψ)

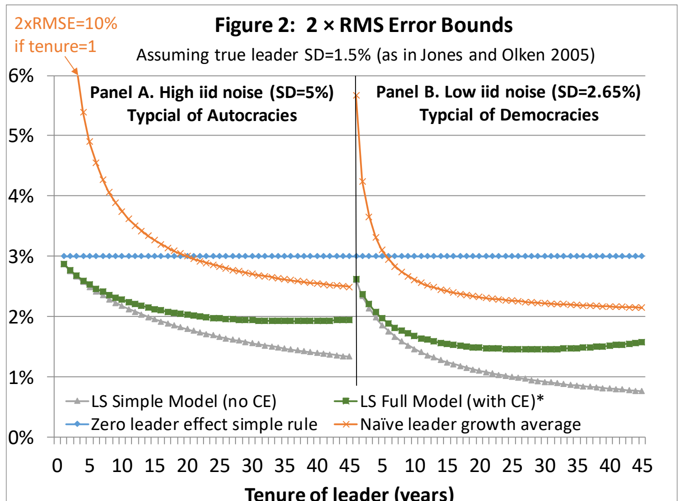
Assuming true leader SD=1.5% (as in Jones and Olken 2005)



Notes: Simple Model has no country effects (SD country effect=0).
* Full model calculation assumes SD country effect=1%, 60 yrs of data per country, average other leader tenure 6yrs.

Figure 2: 2 × RMS Error Bounds

Assuming true leader SD=1.5% (as in Jones and Olken 2005)



Notes: Simple Model has no country effects (SD country effect=0). Root Mean Squared Error assuming A1 and A2 as in main text. * Full model calc assumes SD country effect=1%, 60 yrs of data per country, average other leader tenure 6yrs.

Also striking is the degree of imprecision of estimates of leader effects for autocrats, even for leaders with a long tenure and using a least-squares methodology. Specifically, the Least Squares 2x RMSE bounds for autocracies never get much below 2%, suggesting that a leader with 4% per capita growth (2% above the cross-country mean) will *never* be statistically significant (using our definition above) – regardless of how long he or she is in office. The figure also highlights how much wider are the SE bounds in autocracies than democracies, and hence the greater is the need to avoid spuriously attributing successful growth to leaders of autocratic countries.

These graphs show how long tenures have an ambiguous effect. On one hand, longer tenures even out iid noise. In a simplified model (no country effects $\sigma_c^2 = 0$), this means that shrinkage factors keep rising, and RMSEs keep falling with tenure) On the other hand, a long tenure also means there is less data available to distinguish the leader effect from the country effect (for fixed N, the length of the dataset). So in the full model plots of shrinkage factors and errors are *backward bending*: at some point, an extra year of tenure obfuscates the measurement of country effects to the point that lines are flat, and after that an extra year of tenure actually *reduces* the shrinkage factor and *increases* errors. This suggests that even autocratic leaders with very long tenures will not have shrinkage factors close to one.

Section 2.3: Estimates of true population variance components

In order to produce our own least-squares estimate of leader *i* on growth $\hat{\mu}_i = \hat{\psi}(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})$ we need to calculate $\hat{\gamma}$ and $\hat{\psi}$, which depend on estimates of the variance components σ_c^2, σ_μ^2, σ_ε^2.[22]

Estimating the size of the leader effect (σ_μ)

The difficulty of estimating σ_μ^2 has been long recognized in the random effects panel literature, where estimates of σ_μ^2 and σ_ε^2 are needed to perform Generalized Least Squares. Baltagi (2005 p16) shows that $\hat{\sigma}_\mu^2$ can be backed out from the estimates $\hat{\sigma}_\mu^2 = \hat{\sigma}_{\mu 1}^2 - \hat{\sigma}_\varepsilon^2 / T$ using Equation (17) where $\hat{\sigma}_\varepsilon^2 = \frac{1}{N_L(T-1)} \sum_{i-1}^{N_L} \sum_{t=1}^{T} (g_{it} - \bar{g}_i)^2$ and $\hat{\sigma}_{\mu 1}^2 = \left[\frac{1}{N} \sum_{i=1}^{N} (\bar{g}_i - \bar{g})^2 \right]$ are estimated using standard variance formulas (here, formulas are provided for balanced panels).

$$\hat{\sigma}_\mu^2 = \hat{\sigma}_{\mu 1}^2 - \hat{\sigma}_\varepsilon^2 / T \tag{17}$$

It is possible for $\hat{\sigma}_\mu^2$ to be negative if $\hat{\sigma}_{\mu 1}^2$ is small and so the estimator replaces negative estimates with zero (i.e. $\hat{\sigma}_\mu^2 = \max(0, \hat{\sigma}_{\mu 1}^2 - \hat{\sigma}_\varepsilon^2 / T)$), with the Monte Carlo studies finding this not being a serious problem (Baltagi 2005 p18). Our main estimates use a similar method but with a small-sample adjustment for unbalanced panels from Baltagi and Chang (1994) (which we label SA, after the Stata random effect option that implements it) – as leader tenures are extremely unbalanced. In the appendix, we also report results using a standard random effects estimator without the small sample adjustment (called RE). Monte Carlo simulations reveal that the SA method isn't more accurate on average (less biased) than the RE method, but is substantially less variable across repetitions.[23]

Estimating the size of the other variance components (σ_ε^2 and σ_c^2)

We also need to estimate the size of the iid error and the country effect in order to calculate the least-squares leader estimate.[24] The iid error is allowed to vary across countries, and is estimated using a SA random effects regression country-by-country. Monte Carlo simulations below reveal that it is estimated accurately. The country effect is estimated by adding country dummy variables to the random effects regression of growth residuals, and then calculating their variance. In principle, this has the same upwards bias problem as estimating the variance of leader effects, due to the averaging of the iid error and leader effects. However in practice the average sample length for a country is around 10 times that for a leader, and so the size of the bias is much smaller (we verify this via Monte Carlo simulations).

[22] σ_μ^2 is of interest, because it measures how much leaders affect growth in general. Intuitively, if growth changes a lot *between* leaders then σ_μ^2 will be large, whereas if there is a lot of variation in growth *within* leader terms, then σ_ε^2 should be large.

[23] Note that one can't just calculate the variance of leader growth average $\sigma_{\mu 1}^2$ as this would be substantially upward biased. The xtreg, fe command in Stata reports exactly the unadjusted variance of fixed effects, and so has the same upward bias.

[24] In earlier drafts we also controlled for serial correlation in the error, but it turned out to be difficult to estimate and distinguish from country effects.

Section 2.4 Relation to Teacher Value Added Literature

The teacher value added literature seeks to answer a similar question to the one we ask here with leaders replacing teachers and growth replacing test scores: suppose a particular teacher (or leader) were randomly replaced, how much would test scores (economic growth) increase? The teacher VA approach is promising because the methods in that literature have generally been quite successful.[25] 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.

Our approach here follows the methodology in the teacher value added literature exactly in special cases (Kane and Staiger (2008), and Chetty et al (2014) among others). For example, abstracting from country effects ($\sigma_c^2 = 0$) and classroom effects ($\sigma_\theta^2 = 0$) and when the teacher only teaches one student per class ($n = 1$), the shrinkage factor in Equation 7 here is the same as Equation 9 in Chetty et al (2014) (where the student error corresponds to annual GDP growth error).[26] However, there are also important differences such as our smaller sample size, the additional country effect and our reduced ability to control for exogenous determinants of growth.

A key methodological contribution here is calculating the precision of the individual leader effects and the set of "statistically significant" leaders based on Bayesian credible intervals. Discussions of precision and significance of individual teacher's VA are rare in that literature.

Section 3: Data

Section 3.1: Data Sources

In order to estimate the size of leader effects we need data on leaders, economic growth and a measure of whether each country is an Established Democracy (DEM) or Autocracy/Transition Country (AUT). Data on which leaders are in power is taken from Archigos 4.1 dataset (Goemans et al 2009). In cases where there are multiple leaders in office in a year in same country, we allocate the year to the leader in power for the most days.[27]

Our main source of data on real GDP per capita growth the Penn World Tables (PWT) version 9 (the latest version at the time of writing, Feenstra et al 2015). We also use real per capita GDP growth from PWT 7.1 (Heston et al 2012) and the World Bank's World Development Indicators (WDI) as a cross-check, with results mostly presented in the appendix.[28] We prefer the PWT9 dataset because it has the longest sample (relative to WDI) and an updated methodology (relative to PWT 7.1) for calculating growth rates based on real national prices. Following Jones and Olken (2005), we use the log growth rate: $\ln(Y_t)$- $\ln(Y_{t-1})$, where Y_t is real GDP per capita. Countries with less than 30 years of growth data are dropped, which affects mostly ex-communist countries.[29] See the Appendix for further details on data sources and construction.

---

[25] These methods are not without criticism. For example, Rothstein (2010) finds that some of assumptions of VA models are violated which can lead to future teachers affecting past test scores, and that teacher VA estimates fade out quickly.

[26] Equation 9 in Chetty et al is the same as Equation 5 in Kane and Staiger (2008) with a constant classroom size.

[27] In a previous version of this paper, we used leader data from Jones and Olken (2005). However, that leader data finished in 2000 which meant excluding all leaders from the past ≈ 15 years from our sample, whereas Archigos data runs until 2015.A previous version of the paper allocated years to the leader on 1 January, which produced generally similar results.

[28] In a previous version of this paper we also used PWT 6.1 (as used by Jones and Olken 2005) and also the Maddison dataset (as used by Besley et al).

[29] Dropping ex-communist countries is also beneficial as these countries suffered big falls in output after the communist transition – and sometimes large bounceback – a dynamic that would have to otherwise have to be controlled for.

Regional definitions (for region X year fixed effects) are taken from the World Bank, though with some modifications. The five regions are: Sub-Saharan Africa (SSA); Middle-East and North Africa (MENA); Latin American and the Caribbean (LAC); Asia (a combination of World Bank regions of East Asia & Pacific (EAP) and South Asia, due to the small number of countries in the latter). The modification to World Bank regions is to add to the Europe and Central Asia (ECA) the European offshoots USA, Canada, Australia and New Zealand. ECA is dominated by European countries in our sample due to exclusion of former Soviet Republics and Eastern European countries with less than 30 years of data. Combining the European offshoots with Europe has been done previously in the literature as the European offshoots are close to the frontier, and so are unable to experience fast "catch up" growth. This region is similar to the widely used region of OECD countries (original members). All of the regions have around 20-25 countries, except SSA which has around 40.

Established democracies are defined as countries with an average Polity IV score >.7.5. This is somewhat stricter than the Polity>0 score used by Jones and Olken (2005), and consists almost entirely of countries with a long history of democracy -- which is why we used the nomenclature "established democracies" (abbreviated to DEM). However, it is only slightly stricter than the 6-10 range for democracies recommended in the Polity IV documentation (anocracies are -5 to 5 and autocracies are -10 to -6). We choose an average value rather than reassessing each country's status year-by-year to minimize transitions in and out of democracy, and we use a value above 6 to make sure that democracies did not spend much of the sample as non-democracies. Other countries with an average Polity IV score <7.5 are either autocracies, or transition countries who spent much of the sample as autocracies or anocracies – even if they are democracies today. For simplicity we label leaders of autocracies/transition countries as autocrats (abbreviated to AUT). Several small countries without a polity score are dropped.

Section 3.2: Descriptive Statistics

We have 136 countries for which we have growth, leader and polity data, of these about 20% are established democracies (see the Appendix a full listing). The sample is 1951-2014 for PWT9 growth data, 1951-2010 for PWT7.1 growth data and 1961-2014 for WDI growth data.[30]

Table 1: Descriptive Statistics

Growth Dataset	Sample	Mean	SD	Obs	Leaders	Ave. Tenure
PWT9 (main dataset)	All	1.91%	6.28%	7214	1141	6.3
	Aut	1.77%	6.91%	5598	806	6.9
	Dem	2.40%	3.20%	1616	335	4.8
PWT7.1	All	1.89%	6.81%	6764	1089	6.2
(appendix/	Aut	1.75%	7.45%	5272	776	6.8
robustness)	Dem	2.42%	3.72%	1492	313	4.8
WB WDI	All	1.86%	5.87%	6275	998	6.3
(appendix/	Aut	1.76%	6.43%	4966	734	6.8
robustness)	Dem	2.24%	2.90%	1309	264	5.0

Table 1 shows basic descriptive statistics. We have around 7000 observations and 1100 leaders. Average per capita growth is about 1.9% per annum, and is higher on average in established democracies than autocracies/transition countries. The unconditional variance of growth is much higher for AUT countries than DEM ones.

[30] WDI growth data is available for 2015 and 2016 (for some countries), though we choose to finish the sample in 2014 for comparability with PWT9 and also because more recent data are more likely to be revised by statistical agencies.

Section 3.3 Outliers

Per capita GDP growth rates are often very volatile and a small number of observations can have a large effect on estimated results. Intuitively, this is because the importance of the observation increases with the square of its size: a growth rate 50ppts above the mean has 10000 times the weight of one 0.5ppts above the mean. These extreme observations do exist, for example, for countries entering or exiting civil wars. As such, a couple of coincidental leader transitions around times of civil wars or other extraordinary events can substantially influence our results.

We take a very conservative definition of outliers – log growth of more than 40% (in absolute value) in a particular year – and drop these from our main results. There are only around 17 outliers for PWT9 data by this criterion, which is around 0.2% of the 7000 observations (listed in Appendix Table 1B). We also drop Kuwait in 1990-91 as in these years Kuwait was occupied by Iraq and so was not a separate country (growth was also unsurprisingly volatile). We also drop Liberia as it is an extremely influential country – excluding that country shifts the leader SD by around 0.2% which is around one standard error (due to several large outliers at periods of leader transitions). Finally we exclude Myanmar before (and including) 1970, because of a three-fold increase in GDP per capita that occurred that year (probably due to a change in the methodology for calculating GDP).

Two aspects of the outliers are striking. First is the number of extreme observations that coincide with wars. Some of the largest outliers include in Iraq during the Gulf War, the Rwandan genocide of 1994, the Lebanese civil war in the 1970/1980s. The second striking fact is the level of disagreement about growth rates during these periods: the average difference between the maximum and minimum growth rates in each year across the three datasets (PWT9, PWT7.1 and WDI) is 23% (for the individual years)! This reflects the difficulty of measuring the change in per capita output during extreme times like civil war or genocide, and further justifies dropping the most extreme values from the dataset.

Section 4: Monte-Carlo Results

Table 2 tests the performance of our least-squares methodology and alternative simple rules using Monte Carlo simulations where growth components $\{\mu_i, \mu_c, \varepsilon_t\}$ drawn from normal distributions are combined with the actual leader structure of the Archigos dataset.[31] As we draw and observe the true leader effect (μ_i), we can compare it to the leader estimate $\hat{\mu}_i$ to get a sense of the forecast bias ($\lambda \neq 1$) and size of the root mean squared error (RMSE $= \sqrt{L^{-1}\sum_i(\mu_i - \hat{\mu}_i)^2}$) for the least squares methodology and simple rules. For now, we assume the *same* underlying variation in leader quality as what Jones and Olken (2005) found, $\sigma_\mu = 1.5\%$, so this section also provides guidance on how to calculate individual leader effects taking JO's variation in leader quality as given. In the next section, we will do our own estimate of σ_μ (which turns out to be quite similar).

Conditional on Equation 2 being the true model of growth, there are two possible sources of error in our least squares leader estimate: (i) a method that is either inherently inaccurate (or produces biased forecasts), or (ii) errors in our estimates of the variance components $\{\sigma_u^2, \sigma_c^2, \sigma_e^2\}$ used to construct ψ, γ and the least squares estimates of leader effects. The first row of Table 2, Panel A tests the first hypothesis by assuming we know the true variance components $\sigma_u = 1.5\%, \sigma_c = 1\%, \sigma_e = 5\%$ (close to what we estimate below), which removes errors caused by their estimation. One can see that -- as expected from Proposition 1 -- the least squared method is unbiased ($\lambda \approx 1$), and its RMSE is the lowest of any method in the table (Column C). Note, however that the RMSE of 1.28% is not

[31] That is, the actual PWT9 growth residuals are replaced by a leader effect (μ_i), a country effect (μ_c), and an iid error (ε_t), each drawn 500 times from a normal distribution with mean zero. The leader structure of Archigos remains unchanged. We don't model observables like region X yr fixed effects, which is equivalent to starting with a growth residual.

that much lower than the simple rule of just assuming each leader effect is zero ($\text{RMSE}_{\text{zero}} = 1.5\%$). That is, even with perfect information on $\{\sigma_u^2, \sigma_c^2, \sigma_e^2\}$, the least squares methodology only buys a reduction RMSE of 0.22ppt for the typical leader.

Estimates of the variance components (when σs are unknown) are shown in Column A of Table 2 (estimated using the random effects estimator SA with an unbalanced panel adjustment, as described in Section 2.3). Unfortunately, one can see that the estimate of the variation in underlying leader quality σ_u is substantially upward biased; 1.76% rather than 1.5%, and this gap is more than two standard deviations across Monte Carlo replications. As we show in the Appendix, this is mostly due to the estimation of country effects via country dummy variables - when there are no country effects (and no dummy variables), estimates of σ_u are very accurate using the SA method (Appendix Table 2).[32] While the RE method without an unbalanced panel adjustment is more accurate at estimating σ_u , it is also more noisy across repetitions, and sometimes is severely downward biased (delivering an estimate of zero).[33]. Hence, we view the SA method as a more conservative choice when questioning the size of leader effects. The country effects are also substantially upward biased (σ_c of 1.4% rather than 1%) which is due to using the variance of country dummies to approximate CEs (other methods deliver similar results).[34] The estimates of σ_e (SD iid noise) , on the other hand, are estimated very accurately.

Table 2: Monte Carlo Estimates (500 reps)

	True: sd(leader)=1.5%; sd(country)=1%; sd(iid)=5%						
	A. Estimated Variance components			B. Shrinkage Coeff (across leaders)		C. Performance	
	SD(leader)	sd(iid)	sd(CE)	ψ (shrinkage)	γ (country adj)	Bias (λ≠1)	RMSE (SE)
	Panel (1): Optimal Least Squares leader Estimates (full model)						
When σ are known	1.50%	5.00%	1.00%	0.28	0.57	0.99	1.28%
	(known)	(known)	(known)	(0.15)	(0.09)	[0.04]	[0.03%]
When σ are unknown	1.76%	5.00%	1.40%	0.33	0.69	0.84	1.29%
(need to be estimated)	[0.11%]	[0.05%]	[0.08%]	(0.16)	(0.09)	[0.08]	[0.03%]
	Panel (2): Simple Rules						
Naïve leader growth ave	-	-	-	1	0	0.18	3.21%
Zero leader Effect				0	Any	-	1.50%

* Note: Table presents Monte Carlo estimates of (A) variance components, (B) the shrinkage coefficients used to construct least squares leader estimates and (C) measures of the performance of these estimates. Here the actual country X leader tenure structure is used, but the growth process is drawn from a normal distribution. A successful method uncovers the "true" parameter of the leader effect of 1.5%, CE of 1% and iid SE of 5% (Panel A) but more importanly is unbiased (λ=1) and has a lowest root mean squared errors (also known as the standard error) (Panel C). Panel B presents the shrinkage coefficient ψ and γ used to produce the leader estimates (mean and standard deviation across leaders). Calculations of variance components use the "SA" method a described in the text. Standard deviations across Monte Carlo replications are reported in brackets and standard deviations across leaders are reported in parentheses. Results are similar using alternative assumptions and methods, as described in the appendix.

The implications of estimating the variance components can be seen comparing the performance indicators across the first two rows of Panel 1 of Table 2. Despite the estimation noise and biased estimates of σ_u and σ_c , the costs in terms of accuracy are low: the RMSE only increases from 1.28% to 1.29% (which reduces our concern about the bias in $\hat{\sigma}_\mu$ and $\hat{\sigma}_c$). However, the upward bias in σ_u and σ_c leads to upward bias in the measured shrinkage factors

[32] That is, the model finds it hard to distinguish between a good country effect and a string of good leaders. Computationally we estimate the country effects first, and the leader effects are deviations form a country average.

[33] We show RE does particularly badly for example when there are heteroskedastic leptokurtic errors in Appendix Table 2

[34] One potential solution for more accurate estimation of the variance components σ_u and σ_c is an indirect inference approach: choosing a σ_u , σ_c such that the biased estimate from the MC matches the biased estimate in the data. This is an interesting area for future research.

ψ and γ. Specifically, a higher estimate of σ_u increases the numerator of $\hat{\psi}$ in Equation 7, which increases the average shrinkage factor across leaders from 0.28 (with full information) to 0.33. Similarly, a higher estimate of σ_c increases the average value of $\hat{\gamma}$ from 0.57 (with full information) to 0.69 (Eq 8). Combined, this means that the least squares leader effect $\hat{\mu}_i$ varies *too much* across leaders relative to the variation of true leader quality, and so a 1ppt increase in $\hat{\mu}_i$ is only associated with a 0.84ppt increase in the true leader effect. That is, the point estimate of bias is $\hat{\lambda} = 0.84 < 1$. An average shrinkage factor of $\psi \approx 0.3$ means that a leader growth (residual) of 1% will become 0.3% after removing noise, which is similar to that of an autocracy/transition leader who has been in power for 5 years in Figure 1.[35]

Panel (2) presents the performance of our two simple rules: assuming zero leader effects or naively using the leader growth average. The leader growth average performs very poorly: it is severely forecast biased ($\lambda = 0.18$), such that a 1ppt increase in the leader growth average is only associated with a 0.2 ppt increase true leader quality. Worse, it has huge RMSE, which is around 2.5 times that of the least squares methodology for a typical leader. However, the zero leader effect performs surprisingly well: the RMSE is less than half that of the leader growth average and only around 0.2ppts higher than the least squares method (forecast bias is not defined). This suggest that even with Jones and Olken's estimate of the underlying variation of leader quality ($\sigma_\mu = 1.5\%$), allocating the leader growth average to a leader is very inaccurate, and policymakers are usually better off just assuming a leader growth average of zero.

Section 5: Estimates of variance components in the data

We now begin estimating leader variance components using the real data. These are not the *observed* least-squares leader effects that we can calculate for each leader in the data, but rather standard deviations of *unobserved* variables μ_i, μ_c, ε_{ict} in the variance components model of growth in Equations (1) and (2). The variance components calculated in this section are mostly of interest as a building block for the calculation of the shrinkage parameter and least squares leader estimates in the next section, though they are also important indicators of leader quality in their own right. However, in the real world, observers are usually trying to infer whether a particular leader is good for growth, which is why we focus on identifying leader quality of individual leaders.

Table 3: Estimates of Variance Components

	Pooled			Autocracies/Transition			Established Democracies		
	sigmaU	sigmaC	sigmaE*	sigmaU	sigmaC	sigmaE*	sigmaU	sigmaC	sigmaE*
	Panel A: Estimates of variance components (from data)								
Actual Estimates	1.69%	1.27%	4.60%	1.96%	1.39%	5.01%	1.11%	0.72%	2.65%
	Panel B: Country-level block bootstrap (1000 reps)								
BS Mean	1.68%	1.26%	4.61%	1.96%	1.37%	5.01%	1.08%	0.70%	2.65%
BS Median	1.68%	1.26%	4.60%	1.96%	1.37%	5.01%	1.11%	0.70%	2.65%
SD (across BS reps)	0.20%	0.11%	0.20%	0.22%	0.13%	0.23%	0.29%	0.14%	0.26%
Lower 95% CI	1.27%	1.04%	4.19%	1.52%	1.11%	4.56%	0.47%	0.43%	2.12%
Upper 95% CI	2.05%	1.49%	4.99%	2.36%	1.63%	5.45%	1.59%	0.97%	3.16%
Panel C: Two tailed BS p-value equality of variance comp. (Aut/Trans = Dem):							1.40%	0.00%	0.00%

Notes: Panel A reports estimates of variance components using the SA method on PWT9 data (excluding outliers), Panel B reports how these change across bootstrap samples where countries are sampled at random (with replacement). Panel C: The fraction of bootstrap reps (x2) where the variance larger for DEM. * SigmaE is the across-sample SD of the iid error. In the estimation of leader effects, we allow sigmaE to vary across countries.

[35] When we estimate separately for autocracies and democracies, with leptokurtic errors $\hat{\psi}_{AUT/TR} = 0.33$ and $\hat{\psi}_{DEM} = 0.46$.

Using the unbalanced panel adjustment method (SA), the unobserved leader component is estimated to have SD of 1.7% ($\hat{\sigma}_{\mu}$) across all leaders (Table 3 Panel A), with a 95% confidence interval of [1.3, 2.0] based on the bootstrap standard errors (Panel B). Other datasets and methodologies generally produce smaller estimates of overall and autocratic underlying leader SD (see Appendix Table 3 and Section 8), and so we view our baseline modelling choices as favorable to the "leaders matter for growth" hypothesis.[36] Moreover, our estimate of the SD of underlying leader quality is slightly larger than the baseline estimate in JO of 1.5%.

We find that autocratic/transition countries have a higher SD of leader quality (2.0%) than do established democracies (1.1%). Using bootstrap standard errors, we can reject at the 5% significance level that the variation of leader quality is the same in the two groups.

As previewed earlier, the year-to-year growth variation (the iid error component) is much noisier in autocracies than democracies, which is important for identifying the size of the least-squares leader estimate in the next section. The standard deviation of iid error is a remarkably high 5.0% for autocratic/transition countries, compared to 2.7% for established democracies. A plausible explanation for this is that autocratic/transition countries tend to be lower income countries than established democracies, and are more vulnerable to shocks like natural disasters and commodity prices. It is also likely lower income countries have less administrative machinery for collecting GDP data, leading to higher measurement error. This noise plays a major role in estimating leader effects. In the calculations of individual least squares leader estimates in the next section, we estimate the iid error standard deviations separately for each country.

Country effects are also larger for autocracies/transition countries than established democracies. The estimated SD of the country effect is around 1.4% for autocracies/transition countries, whereas for established democracies the country effects are around 0.7%. Thus all three sources of growth variation in autocracies/transition countries are higher than in established democracies, only one of which (and not the largest) is leader quality.

Section 6: Least squares leader effects - estimates and scatter plots

Now that we have estimates of the variance components (from Table 3), we can use them to produce estimates of the least squares leader effect $\hat{\mu}_i = \hat{\psi}(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})$ for every national leader in our dataset using Equations (7) and (8). The variance components used to construct $\hat{\psi}_i$ and $\hat{\gamma}_{\mathrm{i}}$ for each leader use the underlying SD of the leader effect ($\sigma_u^{AUT} = 1.96\%$, $\sigma_u^{DEM} = 1.11\%$) and the SD country effect ($\sigma_C^{AUT} = 1.39\%$, $\sigma_C^{DEM} = 0.72\%$), estimated separately for autocratic/transition countries and established democracies as in Table 3 Panel A, with the SD of the iid noise (σ_e) estimated country-by-country.[37]

The relation between the raw average growth rate under each leader (x-axis) and the least-squared leader effect (y-axis), is shown in Figure 3, Panel A (each leader is a dot). If the average growth rate under each leader was

[36] Specifically, using the RE methodology produced substantially lower estimates of leader effects for all types of leaders, though the effects are the most dramatic for all leaders and autocracies, where in the leader SD could halve or even go to zero (for example RE method and PWT 7.1 data). Estimates of iid noise and country effects are similar across methods. See Section 8 for a discussion of different datasets.

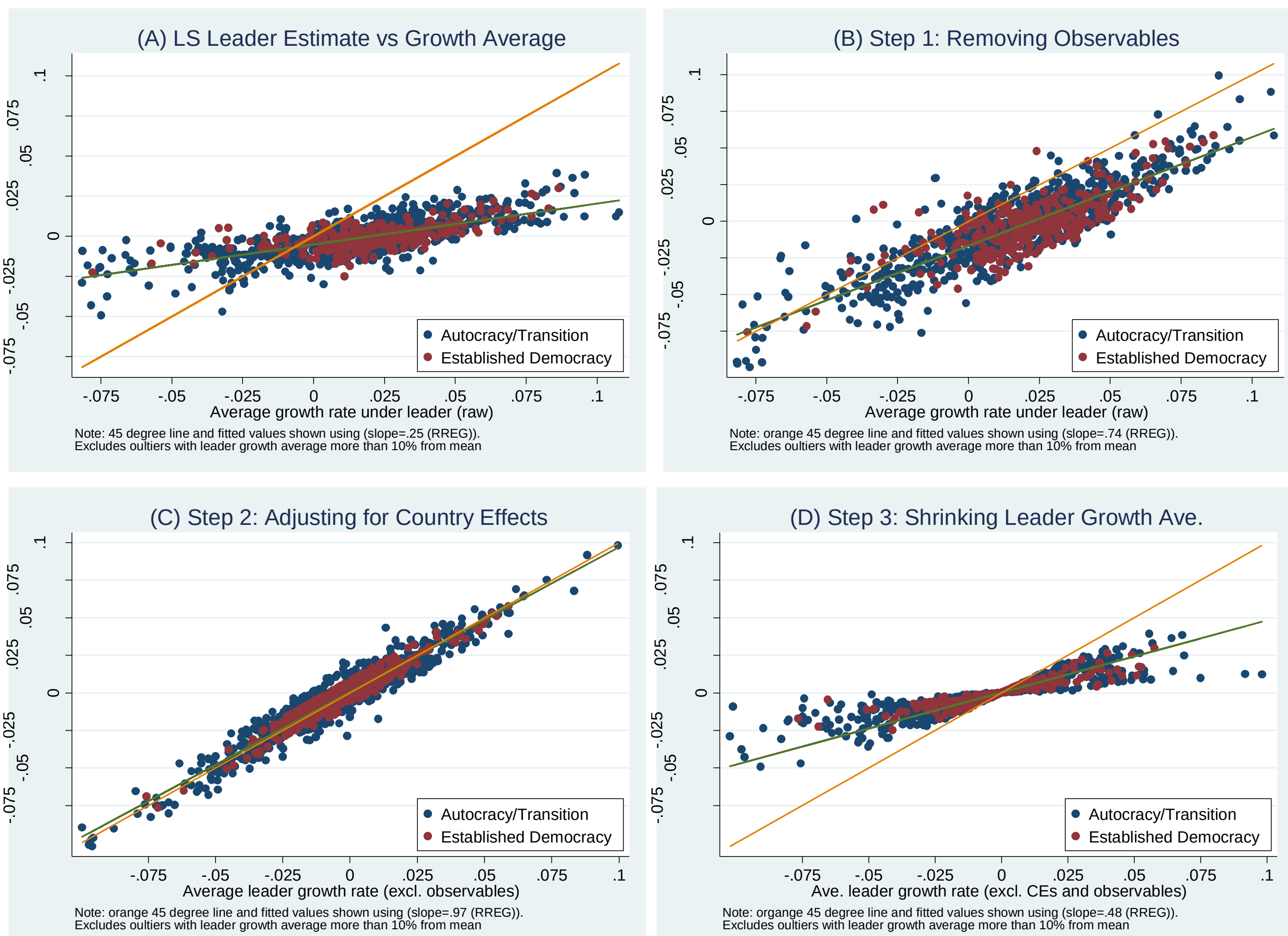
[37] Appendix Figure 2 provides plots of how the shrinkage factor (ψ), errors, and country adjustment factor (γ), change with tenure and country-level iid noise, for autocracies/transition countries and democracies. A short tenure and small $\psi < 0.2$ are common among autocratic leaders (which is not visually obvious in the figure as all these leader estimates are on top of each other). However a small ψ is rare in established democracies.

perfectly informative about that leader's true quality, the dots representing leaders would line up exactly on the 45 degree diagonal (slope of 1). If they were completely uninformative about leader quality, the fitted line would be horizontal. Although the fitted line is slightly upward sloping (slope=0.25), it is much flatter than the diagonal. This suggests that on average, a 1ppt increase in the raw leader growth average is associated with a 0.25 ppts increase in the LS leader effect.

The path of adjustment from raw leader growth average to least-squares leader estimate involves 3 steps, which are shown in remaining three subplots in Figure 3. Figure 3B shows the effect of removing observables (region x year FE, and the cross-country growth mean) which reduces the size of leader effects modestly – a slope of 0.75 means that a 1ppt increase in the raw leader growth average is associated with a 0.75ppt increase in the leader growth average after adjusting for observables. This adjustment also makes growth rates mean zero.

Adjusting for country effects (by subtracting $\hat{\gamma}\overline{g}_{-ic}$) does not reduce the leader growth average substantially in general, though can be important for specific leaders (as we see in the next section). Figure 3C plots the leader growth average after removing observables (x-axis) vs leader growth average adjusted for country effects and observables (y-axis).[38] The two are highly correlated, with a slope of 0.97, which is close to the 45 degree line.

Figure 3



[38] Note that the *adjusted* leader growth average can be higher or lower than the raw leader growth average, depending on whether other leaders had low or high average growth.

The final step of adjustment – and the most important --- is "shrinking" leader effects. Figure 3D plots the *adjusted* leader growth average on the x-axis (after removing observables and country effects) vs the least squared leader estimate on the y-axis, with the ratio of the two representing the shrinkage factor ψ. On average a 1ppt increase in the adjusted leader growth average only leads to an increase in the LS leader effect by 0.5.

Calculating statistically significant leaders

Figure 4 plots the combination of LS leader effects and the size of Root Mean Squared Errors (RMSE) for autocracies and democracies. The set of significantly good-for-growth leaders are in the bottom right corner, with a large leader effect and relatively small errors. The worst leaders are in the bottom left corner, with strongly negative leader effect and relatively small errors. The other leaders have errors too large relative to their leader effect (center region), which lead to a leader estimate that is insignificantly different from zero. Two features are illustrated by Figure 4. First, the vast majority of leaders have an insignificant leader effect (they are in the center region), with only a small fraction of leaders falling on either side. Second, the RMSE of democrats (generally <1%) is much smaller than that of autocrats (generally between 1-2%). As such, even though many autocrats have more positive or negative LS effects, they also have larger errors, and so still fall in the insignificant center region.

Section 7: Who are the best and worst leaders?

We define the best leaders as any leaders for whom the estimated leader effect is positive and significant at the 95% level. The worst are those with negative and significant leader effects. Table 4 shows the best while Table 5 shows the worst as ranked by their estimated leader effects. Only 47 out of 762 leaders (6 percent) with tenures of at least 3 years have significant positive or negative leader effects (as defined by Definition D2 based on the 95% credible interval).[39] This is not a very impressive showing when significance is defined at the 95 percent level. However, this does not cast doubt on the existence of leader effects in general, since Table 3 already demonstrated a significant result on the variation in leader quality. The low number of significant effects instead shows the difficulty of detecting *which* leaders represent a draw of high or low leader quality amidst the other factors and high noise. There is nothing that guarantees that only a small fraction of leaders have a contribution that is statistically significant - it's a function of the size of leader effects and noise. Appendix Figure 3 shows from Monte Carlo simulations that if growth were less noisy (an iid SD of 1% rather than 5%), then around 1/3 of leaders would be statistically significant.

Reviewing some of the best and worst leaders helps show why our method diverges so much from the prevailing practice of giving the leader credit for all of the raw growth average during his or her tenure.

The region-year adjustments reduce the magnitude of almost all leader estimates. To put it most simply, if the world on average across regions and years is growing about 2 percent per capita (see Table 1- descriptive statistics), then on average we would subtract about 2 percentage points from all adjusted leader average growth rates.

This adjustment will be more or less depending on regional growth averages for the whole sample and the timing of regional business cycles for each leader. Averaged across leaders, per capita growth was almost 3.5% in Asia, around 2.5% in Europe and European offshoots and around 1.5% 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 shared regional culture, history, trade and investment

[39] We focus on leaders with a tenure of 3 years or more because annual growth data introduces substantial rounding errors for leaders of shorter tenures. For example, a leader who entered in June of one year and exited in July in the following year would be allocated two years of tenure (as they had the longest tenure in each individual year), despite being in office for only 13 months. This problem is worse the more leader transitions in each year.

networks, and geography. Because growth under the typical Asian leader was 2 percentage points higher than the typical African leader, the typical Asian leader needs to achieve raw growth 2 percentage points faster than the typical African leader to be allocated the same leader effect. So for example, leaders like Seretse Khama of Botswana and Yoweri Museveni of Uganda outrank Lee Kuan Yew in part because they did well relative to the lower average regional growth in sub-Saharan Africa compared to East and South Asia.

Table 4: Significantly Best Leaders (tenure 3+ yrs)

Rank	Name	Country	LS Leader Est.	RMS Error	Raw Growth Ave	Shrinkage (ψ)	Sig 99%	Tenure	Dem	1stYear
1	Than Shwe	MMR	3.95%	1.06%	8.58%	0.71	1	19	0	1992
2	Khama	BWA	3.83%	1.30%	9.56%	0.56	1	15	0	1966
3	Rodriguez Lara	ECU	3.63%	1.29%	9.14%	0.57	1	4	0	1972
4	Razak	MYS	3.29%	1.28%	7.47%	0.58	1	5	0	1971
5	Medici	BRA	3.09%	1.12%	8.71%	0.67	1	4	0	1970
6	Ikeda	JPN	2.98%	0.77%	8.65%	0.52	1	4	1	1961
7	Papadopoulos	GRC	2.93%	0.99%	8.21%	0.75	1	6	0	1968
8	Hun Sen	KHM	2.90%	1.13%	5.07%	0.67	0	26	0	1985
9	Chun Doo Hwan	KOR	2.76%	1.03%	8.09%	0.72	1	7	0	1981
10	Sato	JPN	2.65%	0.65%	7.70%	0.66	1	8	1	1965
11	Kishi	JPN	2.54%	0.77%	7.81%	0.52	1	4	1	1957
12	Museveni	UGA	2.42%	0.92%	3.25%	0.78	1	29	0	1986
13	Lee Kuan Yew	SGP	2.38%	1.03%	6.26%	0.73	0	30	0	1961
14	Adenauer	DEU	2.22%	0.57%	6.34%	0.73	1	13	1	1951
15	Santer	LUX	2.02%	0.75%	4.72%	0.55	1	10	1	1985
16	Zhivkov	BGR	1.99%	0.91%	5.08%	0.78	0	19	0	1971
17	Raab	AUT	1.98%	0.58%	5.89%	0.72	1	8	1	1953
18	Chiang Ching-Kuo	TWN	1.97%	0.94%	6.82%	0.77	0	10	0	1978
19	Cristiani	SLV	1.94%	0.98%	3.15%	0.75	0	5	0	1989
20	Hee Park	KOR	1.91%	0.82%	6.42%	0.82	0	19	0	1961
21	Karamanlis	GRC	1.81%	0.81%	4.87%	0.83	0	13	0	1956
22	Franco	ESP	1.69%	0.86%	5.42%	0.81	0	25	0	1951
23	Ahern	IRL	1.67%	0.62%	5.32%	0.69	1	11	1	1997
24	Rumor	ITA	1.66%	0.69%	6.85%	0.61	0	4	1	1969

Notes: list of leaders (tenure >=3yrs) which have significantly positive LS leader effect at the 5% level (two-tailed). That is where (LS leader estimate)-Z*(RMS Err)>0 for Z=1.96. Sig 99% is analagous, but where Z=2.576.

But regional growth also varies over time due to common shocks and business cycles (such as the Asian financial crisis and Latin America's lost decade). A relatively unknown leader like Razak in Malaysia also shows up as better than famous Asian leaders like Lee Kuan Yew and Park Chung Hee because growth under Razak was high during a time when regional Asian growth was low. Eisenhower and Nixon surprisingly show up on the worst list in part because growth under them was low compared to high growth at the time in the region of Europe and European offshoots.

The residual after adjustment will then be compared to the same residual for other leaders for the same country. Some surprising results for high positive estimated leader effects on growth are because the other leaders for the same countries were truly disastrous – examples of such relative successes include Hun Sen of Cambodia and (to a lesser extent) Yoweri Museveni of Uganda. The flip side of this will be that some of those disastrous leaders will show up among the worst leaders in Table 5: Lon Nol of Cambodia and Idi Amin of Uganda.

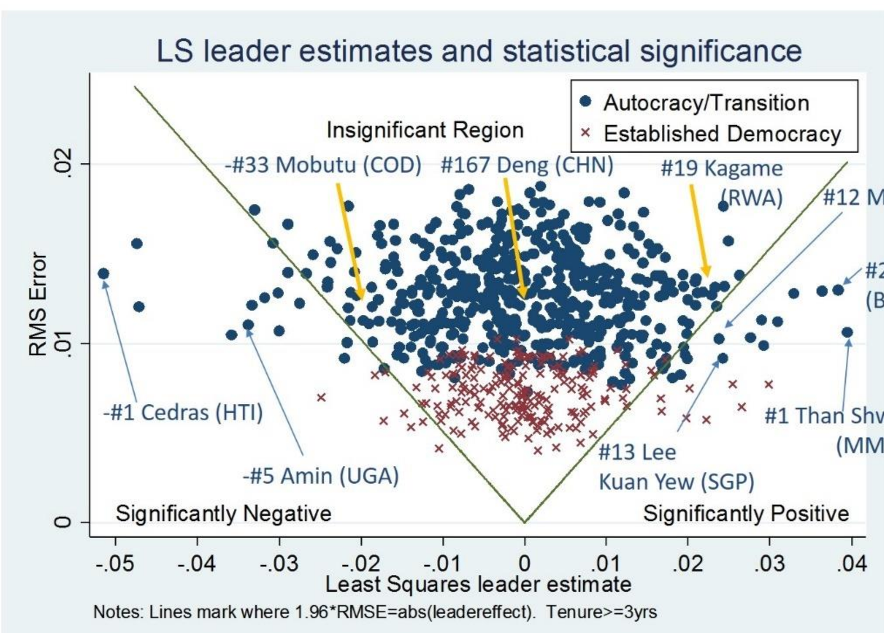
Other less surprising top leaders are those who have unusually high growth rates even for a high growth country, such as Chun Doo Hwan, Lee Kwan Yew, Park Chung Hee, and Chiang Ching-Kuo.

Table 5: Significantly Worst Leaders (tenure 3+ yrs)										
Rank	**Name**	**Country**	**LS Leader Est.**	**RMS Error**	**Raw Growth Ave**	**Shrinkage (ψ)**	**Sig 99%**	**Tenure**	**Dem**	**1stYear**
1	Cedras	HTI	-5.15%	1.39%	-9.85%	0.50	-1	3	0	1992
2	Lon Nol	KHM	-4.74%	1.56%	-11.12%	0.37	-1	4	0	1971
3	Ochirbat	MNG	-4.71%	1.20%	-3.21%	0.62	-1	7	0	1990
4	Antall	HUN	-3.58%	1.05%	-4.87%	0.72	-1	4	0	1990
5	Amin	UGA	-3.38%	1.10%	-2.97%	0.69	-1	8	0	1971
6	Mello	BRA	-3.33%	1.21%	-2.97%	0.62	-1	3	0	1990
7	Gouled Aptidon	DJI	-3.18%	1.25%	-4.28%	0.59	0	22	0	1977
8	Alia	ALB	-3.07%	1.56%	-5.81%	0.37	0	7	0	1985
9	Bandaranaike, S.W.F	LKA	-3.01%	1.28%	-2.85%	0.57	0	4	0	1956
10	Ne Win	MMR	-3.00%	1.07%	0.36%	0.70	-1	18	0	1971
11	Plaek Pibulsongkram	THA	-2.89%	1.40%	-2.89%	0.49	0	7	0	1951
12	Khalifah Ath-Thani	QAT	-2.74%	1.22%	-3.19%	0.61	0	23	0	1972
13	Junichiro Koizumi	JPN	-2.48%	0.70%	1.08%	0.61	-1	6	1	2001
14	Arbenz Guzman	GTM	-2.20%	0.91%	-0.99%	0.78	0	4	0	1951
15	Ma Ying-jeou	TWN	-2.17%	1.00%	2.69%	0.74	0	7	0	2008
16	Shinzo Abe	JPN	-1.83%	0.82%	1.28%	0.45	0	3	1	2007
17	Wilson	GBR	-1.72%	0.57%	0.78%	0.74	-1	7	1	1965
18	A. Papandreou	GRC	-1.72%	0.86%	0.68%	0.81	0	10	0	1982
19	B. Cowen	IRL	-1.71%	0.84%	-4.22%	0.43	0	3	1	2008
20	Nixon	USA	-1.48%	0.61%	1.83%	0.70	0	6	1	1969
21	Berlusconi	ITA	-1.39%	0.53%	-0.12%	0.77	-1	10	1	1994
22	Eisenhower	USA	-1.25%	0.56%	1.23%	0.74	0	8	1	1953
23	Chirac	FRA	-1.05%	0.41%	1.72%	0.86	0	12	1	1995

Notes: list of leaders (tenure 3+ yrs) who have sign negative LS leader effect at the 5% level (two-tailed). That is if (LS leader est.)+Z*(RMS Err)<0 for Z=1.96. Sig 99% is analagous, but where Z=2.576. Rank from worst.

It is also instructive to see why some leaders widely believed to be "benevolent autocrats" missed out on a top ranking. The most surprising is Deng Xiaoping (#167) of China, whose leader effect suffered from the high growth of most other Chinese leaders. Another famous leader missing for the same reason is Chiang Kai-Shek in Taiwan (#98), who was no better at producing growth than other Taiwanese leaders.

Figure 4



Similarly, some famous bad leaders fail to show up partly because other leaders in the same country also had poor performance – this includes Joseph Mobutu of DR Congo, the Ayatollah Khomeini of Iran, and Anastasio Somoza Debayle of Nicaragua.

One can see from Figure 4 that Kagame and Mobutu are close to their respective significance cutoffs, whereas Deng Xiaoping is not. Consequently, if we evaluate significance at the 90% level (rather than in at the 95% as in the rest of the paper), Kagame and Mobutu join the best and worst groups (respectively) - see Appendix Tables 4B and 5B. At the 90% level, the fraction of significant leaders increases from around 6% to 11.5%, and naturally that extra 5.5ppt of significant leaders will include a few who are famous amidst a majority of obscure leaders.

Finally, detecting the leader effect requires relatively high signal-to-noise ratio. This happens with some of the top leaders due to long tenures averaging out noise: Hun Sen (25 years), Khama (15), Museveni (28), Park Chung Hee (19), Lee Kuan Yew (30), and Franco (25).[40] It is also important that all of these leaders also had long periods for their country when they were NOT in office, so that their leader effect could be distinguished from a country effect.

Another major factor in determining the leader effect and its statistical significance is country-specific iid noise. Low noise tends to both raise the LS leader estimate and to lower its error as discussed above in Section 2.2. 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 1970-1974 during the "Brazilian Miracle") and Chiang Ching-Kuo in Taiwan.[41]

A low standard deviation of the error term in established democratic countries also allows the detection of leader effects for little-remembered leaders. Japan accounts for three of these – Ikeda, Sato, and Kishi. Similarly, low iid noise means that democratic leaders from Austria, Germany, Luxembourg, Ireland, and Italy show up among the top leaders. Altogether, democratic leaders account for 8 of the 24 top statistically significant leaders.

The low noise in established democracies also helps explain why some 8 out of 23 leaders on the worst significant list are democratic leaders. This is the other part of the explanation of why Eisenhower and Nixon showed up on the worst list.

Conversely, some famously good or bad autocratic leaders miss out on the best and worst significant list because high iid noise results in smaller shrinkage factors (ψ) and larger errors of the leader estimates. Deng Xiaoping suffered also from very high noise in China (standard deviation of the error term more than 6 percentage points).

The high noise not only lowers the LS leader estimate, but also blows up its error. Paul Kagame of Rwanda has an LS leader estimate among the top leaders (2.4%), but had a high error of the leader estimate because Rwanda has very high noise (more than 6 percentage points). Another celebrated African leader, Meles Zenawi in Ethiopia, had only a modest leader estimate and a high error because of high noise. Likewise, for the worst leaders, Khomeini, Mao, and Somoza have high errors relative to their estimates because of the high noise in Iran, China, and Nicaragua respectively. This rendered their LS leader estimates statistically insignificant.

[40] Franco's tenure here only includes his time in office after growth data becomes available in 1951.

[41] Emilio Medici is little remembered for this growth performance partly because the credit for the Brazilian Miracle is often given instead to the Minister of Finance Delfim Netto. This highlights how our assumption that the leader gets all the credit for growth is very favorable to the leader growth hypothesis.

The worst leader (with at least 3 years tenure) is Raoul Cédras, a military general who led the September 1991 coup against Jean-Bertrand Aristide, Haiti's first plausibly democratically elected leader. The coup led to a series of international trade sanctions that crippled Haiti's economy in the following three years when Cédras was the de facto leader, resulting growth of -7%, 7%, and -14% (respectively) according to PWT9 -- Haiti's worst three years of growth on record. So, it is still true that an extremely bad growth outcome can contribute to a significantly negative leader estimate. However, extreme growth is not sufficient. An even worse growth outcome (-16% on average for three years) occurred in Chad under Goukouni Oueddei, for whom we fail to reject a leader effect of zero. The difference is that iid noise is much higher in Chad than in Haiti.

Another surprise is our overall best leader Than Shwe of Myanmar (in office 1992-2011), an international pariah that usually failed to receive recognition as a benevolent autocrat. Than Shwe's least squares leader contribution of 4% is a function of a high average growth rate (8.5%, off a low base of ≈$200pc at the start of his tenure), a long tenure (19 years), not very high iid noise, and low growth under other leaders in the same country -- another Burmese leader, Ne Win (1962-1989), shows up in the worst table. Than Shwe did preside over partial economic liberalization, compared to the "Burmese Way to Socialism" of Ne Win.[42] The lack of recognition of Than Shwe may reflect doubts about the quality of the growth data in Myanmar. We address the issue of data quality in the next section, but for Myanmar there are no truly independent sources of growth data. And again, high growth is not enough by itself in our method to show up as a top leader. Hu Jintao of China also achieved 8 percent per capita growth over a decent tenure (10 years), but received a low and insignificant leader estimate because of high growth under other Chinese leaders and high iid noise in China.

The good news for the benevolent autocrat hypothesis is that we do confirm strong estimated leader effects for some celebrated leaders, such as Seretse Khama, Lee Kuan Yew, Park Chung Hee, and Yoweri Museveni. The news that is not as good is that other celebrated leaders fail to show a strong leader effect, such as Deng Xiaoping, Chiang Kai-Shek, and Paul Kagame, while other uncelebrated leaders show up instead in the top ranks. The news is also bad in the worst leaders table, where a famous disaster like Idi Amin is mixed with many little known leaders.

We conjecture that this pattern is because attention paid to benevolent and malevolent autocrats is usually based on raw growth averages of leaders, while we have seen that estimating the leader effect requires major adjustments to that raw growth average. Although raw growth averages can sometimes contribute to a strong leader effect (like Than Shwe or Cédras), they are not generally a good predictor of our best and worst leaders in Tables 4 and 5. Of the 47 best and worst leaders in our list, just 17 are in the best 25 or worst 25 according to raw growth. Another 12 of our best and worst leaders are not even in the best 100 or worst 100 in raw growth. This is yet more confirmation in the actual data of what we showed above with Monte Carlo simulations in Table 2, that judging leader quality by raw growth averages leads to large errors relative to the optimal least squares estimator of leader effects.[43]

[42] The Economic Freedom of the World publication of the Cato Institute and Fraser Institute shows Myanmar's Economic Freedom Index increasing from 2.7 in 1990 to 4.2 in 2011 on a 1-10 scale. Note that we drop growth data in Myanmar before (and including) 1970, as in PWT9 the level of GDPPC triples in that year – possibly due to a change in methodology.

[43] Another problem is that poor quality growth data might not reflect actual economic progress (and can be also manipulated by rulers). An example of this is that the set of best and worst leaders changes substantially across growth datasets (see Appendix 2). The PWT9 dataset we use in the main text has the largest sample size and is arguably of better quality. However alternative growth numbers – even for the same country-year observations- suggest that measurement error is probably a major contributor to noise.

Section 8: Conclusions

In this paper, we show that shrinking the average growth rate of dictators towards zero is the optimal way to extract their true contribution from a noisy growth signal -- but that very few dictators have contributions large enough to survive the shrinkage.

Despite starting with a model where "leaders matter" for growth, our main finding is that it is surprisingly difficult to confirm statistically significantly positive or negative leader growth effects for individual leaders. That is, knowing leaders matter for growth *in general* is very different from knowing *which* leaders matter for growth. We confirm significant positive or negative leader effects for less than 50 out of around 750 leaders with a tenure of at least 3 years (around 6% of leaders). Many of these are little known, forgotten, or surprising stars or villains. Autocrats are surprisingly under-represented in the set of statistically significant leaders, mostly because autocratic countries also have more noisy growth processes which make it difficult to isolate true leader effects.

Our methodology does confirm significant leader effects for some famous "benevolent autocrats" and influential leaders such as Seretse Khama, Lee Kuan Yew, Park Chung Hee, and Yoweri Museveni, as well as for a famous growth disaster Idi Amin. However, many autocrats that have received much praise for high growth in the development literature fail to show significant effects by our method, including Deng Xiaoping, Chiang Kai-Shek, and Paul Kagame. We also fail to confirm significant growth effects of famous disasters such as Joseph Mobutu and Anastasio Somoza.

We attribute much of this mismatch between our results and prior beliefs to our methodology that adjusts for noise, regional trends, and country effects. However, we also acknowledge it reflects the inherent narrowness of the value-added approach. For example, our leader value added approach would not give enough credit to leaders like Nelson Mandela or Deng Xiaoping who may have permanently changed the nation they led. Combined with our earlier results, this provides another reason against over-reliance on contemporary growth data when evaluating leaders. Studying the lagged effects on leaders on growth, as well as their broader contributions in other realms, would be an interesting area for future research.

References:

Baltagi B. H. (2005) *Econometric Analysis of Panel Data*, Wiley (3rd edition)

Baltagi B. H. and Y. Chang (1994), "Incomplete Panels: A comparative Study of alternative estimators for the unbalanced one-way error component regression model" *Journal of Econometrics* 62, pp 67-89

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)" mineo, University of Chicago

Brown, C. (2017), "Economic Leadership and Growth" SSRN WP https://ssrn.com/abstract=2169149 (December 7, 2017)

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): 2593-2632.

Curtis, DEA (2015) "Development assistance and the lasting legacies of rebellion in Burundi and Rwanda", *Third World Quarterly*, 36:7, 1365-1381

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), pp 459-483

Easterly W. and R. Levine (2016) "The European Origins of Economic Development" *Journal of Economic Growth*

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

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

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

Heston A., R. Summers and B. Aten (2002) "Penn World Table Version 6.1", Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

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 working paper 14607

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 No. 7540

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

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

Appendix 1: Proofs

Proof of Proposition 1 *Least-squares leader estimate*

Part (1A) Derivation of $\hat{\psi}$ and $\hat{\gamma}$:

(i) Preliminaries

Covariance of leader growth average and other leaders' growth average in the same country:

$$E(\bar{g}_{ic}\bar{g}_{-ic}) = \sigma_c^2 \qquad \text{(A1)}$$

The variance of the leader growth average:

$$E(\bar{g}_{ic}^2) = \sigma_c^2 + \sigma_\mu^2 + \sigma_e^2/T_i \qquad \text{(A2)}$$

The variance of other leaders growth average:

$$E(\bar{g}_{-ic}^2) = \sigma_c^2 + \sigma_\mu^2/L_{-i} + \sigma_e^2/(N-T_i) \qquad \text{(A3)}$$

The least squares problem is:

$$\min_{\gamma,\psi} E[\mu_i - \psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})]^2 \qquad \text{(A4)}$$

(ii) derivation of γ:

FOC wrt γ:
$$2E[\mu_i - \psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})][\psi\bar{g}_{-ic}] = 0 \qquad \text{(A5)}$$

In the general case of $\psi \neq 0$ and $\bar{g}_{-ic} \neq 0$:

$$E[\mu_i\bar{g}_{-ic} - \psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})\bar{g}_{-ic}] = 0 \text{(A6)}$$

Using $E[\bar{g}_{-ic}\mu_i] = 0$, the first term disappears, to leave an expression independent of ψ, which can be rearranged to get Equation (7) in the main text

$$\hat{\gamma}_i = \frac{E(\bar{g}_{ic}\bar{g}_{-ic})}{E(\bar{g}_{-ic}^2)} = \frac{\sigma_c^2}{\sigma_c^2 + \frac{\sigma_e^2}{N-T_i} + \frac{\sigma_\mu^2}{L_{-ict}}} \qquad \text{(A7)}$$

(iii) derivation of ψ:

FOC wrt ψ:
$$E[\mu_i(\bar{g}_{ic} - \gamma\bar{g}_{-ic})] = \psi E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2 \qquad \text{(A8)}$$

Now simplifying $E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2$by substituting out for $\gamma E\bar{g}_{-ic}^2$ (using A7):

$$\begin{aligned} E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2 &= E\bar{g}_{ic}^2 + \gamma^2 E\bar{g}_{-ic}^2 - 2\gamma E(\bar{g}_{ic}\bar{g}_{-ic}) \\ &= E\bar{g}_{ic}^2 + \gamma E(\bar{g}_{ic}\bar{g}_{-ic}) - 2\gamma E(\bar{g}_{ic}\bar{g}_{-ic}) \\ &= E\bar{g}_{ic}^2 - \gamma E(\bar{g}_{ic}\bar{g}_{-ic}) \end{aligned}$$

$$= \sigma_c^2(1-\gamma) + \sigma_\mu^2 + \sigma_e^2/T_i \qquad \text{(A9)}$$

This yields the expression for Equation (8) in the main text:

$$\hat{\psi} = \frac{E[\mu_i(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})]}{E((\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})^2)} = \frac{\sigma_\mu^2}{\sigma_c^2(1-\hat{\gamma}) + \sigma_\mu^2 + \frac{\sigma_e^2}{T_i}} \quad \text{(A10)}$$

<u>Part (1B) Forecast unbiasness of least squares leader estimates</u>

Running a regression of $\mu_i = \lambda\hat{\mu}_i = \lambda\psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})$

$$\hat{\lambda} = \frac{Cov(Y,X)}{Var(X)} = \frac{E[\mu_i\psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})]}{E[\psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic})]^2} = \frac{E[\mu_i(\bar{g}_{ic} - \gamma\bar{g}_{-ic})]}{\psi E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2}$$

Now from Equation (A8) substitute $E[\mu_i(\bar{g}_{ic} - \gamma\bar{g}_{-ic})] = \psi E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2$

$$\hat{\lambda} = \frac{\psi E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2}{\psi E(\bar{g}_{ic} - \gamma\bar{g}_{-ic})^2} = 1$$

□

Proof of Proposition 2 ***Precision ("Credible intervals") of Least Squares Leader Estimates***

<u>Part (2A) Derivation of RMSE:</u>

$$\begin{aligned}
E[\hat{\mu}_i - \mu_i]^2 &= E[\psi(\bar{g}_{ic} - \gamma\bar{g}_{-ic}) - \mu_i]^2 \\
&= E[\psi((\mu_i + \mu_c + \textstyle\sum e_{it}/T_i) - \gamma(\sum\mu_{-it}/L_{-i} + \mu_c + \sum e_{-it}/(N - T_i)) - \mu_i]^2 \\
&= E[\mu_i(\psi - 1) + \mu_c\psi(1-\gamma) + \psi(\textstyle\sum e_{it}/T_i) - \gamma\psi(\sum\mu_{-it}/L_{-i} + \sum e_{-it}/(N - T_i))]^2 \\
&= (1-\psi)^2\sigma_\mu^2 + \sigma_c^2\psi^2(1-\gamma)^2 + \psi^2\sigma_e^2/T_i + \gamma^2\psi^2(\sigma_\mu^2/L_{-i} + \sigma_e^2/(N - T_i)) \\
&= (1-\psi)^2\sigma_\mu^2 + \psi^2\left[\sigma_e^2/T_i + \sigma_c^2(1-\gamma) - \gamma\sigma_c^2 + \gamma^2(\sigma_c^2 + \sigma_\mu^2/L_{-i} + \sigma_e^2/(N - T_i))\right] \\
&= (1-\psi)^2\sigma_\mu^2 + \psi^2[\sigma_e^2/T_i + \sigma_c^2(1-\gamma) - \gamma\sigma_c^2 + \gamma^2\sigma_c^2/\gamma] \\
&= (1-\psi)^2\sigma_\mu^2 + \psi^2[\sigma_e^2/T_i + \sigma_c^2(1-\gamma)]
\end{aligned}$$

Where the second last step use there arranged definition of γ: $\sigma_c^2 + \sigma_\mu^2/L_{-i} + \sigma_e^2/(N - T_i) = \sigma_c^2/\gamma$ (from Eq 7)
Hence Equation (9) in the main test is:

$$RMSE(\hat{\mu}_i) \equiv \sqrt{E[\hat{\mu}_i - \mu_i]^2} = \sqrt{(1-\psi)^2\sigma_\mu^2 + \psi^2[\sigma_e^2/T_i + \sigma_c^2(1-\gamma)]}$$

□

<u>Part (2B) Posterior distribution</u>[44]

The posterior distribution $f(\mu_i|\bar{g}_i, \bar{g}_{-i})$ is obtained by first deriving joint distribution $f(\mu_i, \mu_c|\bar{g}_i, \bar{g}_{-i})$, and then averaging over the 'nuisance parameter" μ_c (as in Gelman et al (2005) p63)[45]:

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) = \int_{-\infty}^{\infty} f(\mu_i, \mu_c|\bar{g}_i, \bar{g}_{-i}).\, d\mu_c \qquad \text{(A11)}$$

We can rewrite the leader growth average as $\bar{g}_i = \mu_i + \mu_c + \bar{e}_i$ (where $\bar{e}_i = \sum\varepsilon_{it}/T_i$) such that $\bar{g}_i|\mu_c, \mu_i \sim N(\mu_i + \mu_c, \sigma_e^2/T_i)$, and the other leaders' growth average as $\bar{g}_{-i} = \mu_c + \bar{\mu}_{-i} + \bar{e}_{-i}$ (where $\bar{e}_{-i} = \sum\varepsilon_{-it}/(N - T_i)$ and $\bar{\mu}_{-i} = \sum\mu_{-ic}/L_{-i}$), such that $\bar{g}_{-i}|\mu_c \sim N(\mu_c, \sigma_e^2/(N - T_i) + \sigma_\mu^2/L_{-i})$. Using Bayes rule, the expressions above, and the fact that μ_c and μ_i are independent, the posterior can be written as:[46]

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) \propto f(\mu_i)\int_{-\infty}^{\infty} f(\bar{g}_i|\mu_c\mu_i) f(\bar{g}_{-i}|\mu_c) f(\mu_c)\, d\mu_c \qquad \text{(A12)}$$

---

[44] Note: The derivation of the posterior $f(\mu_i|\bar{g}_i, \bar{g}_{-i})$ is long, so here we includes the key steps but skip much of the algebra.

[45] Gelman A, Carlin J, Stern S, Dunson D, Vehtari A, and D Rubin (2015) "Bayesian Data Analysis", Chapman & Hall/CRC Texts in Statistical Science, 3rd Edition, Kindle Edition

[46] Multiplicative factors that depend on the data $(\bar{g}_i, \bar{g}_{-i})$, but not on the parameters of interest (μ_i, μ_c), such as $f(\bar{g}_i, \bar{g}_{-i})$ (the joint prior of $\bar{g}_i$ and $\bar{g}_{-i}$), can be swept into a constant of proportionality. We follow the convention in Bayesian statistics by using the "proportional to" comparator ∝ which abstracts from these multiplicative constants in the derivation.

Where $f(\mu_i)$ and $f(\mu_c)$ are the pdfs of the priors on $\mu_i \sim N(0, \sigma_\mu^2)$ and $\mu_c \sim N(0, \sigma_c^2)$.
With some algebra (involving completing the square), the pdfs $f(\bar{g}_{-i}|\mu_c)$ and $f(\mu_c)$ can be combined as:

$$f(\bar{g}_{-i}|\mu_c)f(\mu_c) \propto exp\left\{-\frac{(\mu_c - \gamma\bar{g}_{-i})^2}{2[\sigma_e^2/(N-T_i)+\sigma_\mu^2/L_{-i}]\gamma}\right\} \quad \text{(A13)}$$

Where $\gamma = \sigma_c^2/[\sigma_c^2 + \sigma_e^2/(N - T_i) + \sigma_\mu^2/L_{-ict}]$ is as defined in the main text (Equation 7).[47]

We combine $f(\bar{g}_i|\mu_c\mu_i) \propto exp\left\{-\frac{(\bar{g}_i-(\mu_i+\mu_c))^2}{2\sigma_e^2/T_i}\right\}$ and the RHS of Eq (A13), using $\left[\frac{\sigma_e^2}{N-T_i}+\frac{\sigma_\mu^2}{L_{-i}}\right]\gamma = \sigma_c^2(1-\gamma)$ and completing the square by adding and subtracting X_i^2 where $X_i = [\sigma_c^2(1-\gamma)(\bar{g}_i - \mu_i) + (\sigma_e^2/T_i)\gamma\bar{g}_{-i}]/[\sigma_c^2(1-\gamma) + \sigma_e^2/T_i]$

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) \propto exp\left\{-\frac{(\mu_i)^2}{2\sigma_\mu^2}\right\} exp\left\{-\frac{(\mu_i - \bar{g}_i)^2}{2\sigma_e^2/T_i}\right\} exp\left\{\frac{[\sigma_c^2(1-\gamma)(\bar{g}_i - \mu_i) + (\sigma_e^2/T_i)\gamma\bar{g}_{-i}]^2}{2\sigma_c^2(1-\gamma)(\sigma_e^2/T_i)[\sigma_c^2(1-\gamma)+\sigma_e^2/T_i]}\right\}$$
$$\times \int_{-\infty}^{\infty} exp\left\{-\frac{(\mu_c - X_i)^2}{2\sigma_c^2(1-\gamma)(\sigma_e^2/T_i)/[\sigma_c^2(1-\gamma)+\sigma_e^2/T_i]}\right\} d\mu_c \quad \text{(A14)}$$

One can see that the term in the integral is proportional to a normal PDF in μ_c, with mean X_i and variance $\sigma_c^2(1-\gamma)(\sigma_e^2/T_i)/[\sigma_c^2(1-\gamma)+\sigma_e^2/T_i]$. This will integrate out to a constant which does not depend on μ_i.[48]
Hence we can rewrite Equation (A14) as simply:

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) \propto exp\left\{-\frac{(\mu_i)^2}{2\sigma_\mu^2}\right\} exp\left\{-\frac{(\mu_i-\bar{g}_i)^2}{2\sigma_e^2/T_i}\right\} exp\left\{\frac{[\sigma_c^2(1-\gamma)(\bar{g}_i-\mu_i)+(\sigma_e^2/T_i)\gamma\bar{g}_{-i}]^2}{2\sigma_c^2(1-\gamma)(\sigma_e^2/T_i)[\sigma_c^2(1-\gamma)+\sigma_e^2/T_i]}\right\} \quad \text{(A15)}$$

With some more algebra, one can rewrite Equation A15 as:

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) \propto exp\left\{-\frac{(\mu_i)^2}{2\sigma_\mu^2}\right\} exp\left\{-\frac{(\mu_i-(\bar{g}_i-\gamma\bar{g}_{-i}))^2}{2\sigma_c^2(1-\gamma)+\sigma_e^2/T_i}\right\} \quad \text{(A16)}$$

Equation A16 can be further simplified into:

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) \propto exp\left\{-\frac{(\mu_i-\psi(\bar{g}_i-\gamma\bar{g}_{-i}))^2}{2\psi[\sigma_c^2(1-\gamma)+\sigma_e^2/T_i]}\right\} \quad \text{(A17)}$$

The variance in the denominator of Equation A17 can be written as $\psi[\sigma_c^2(1-\gamma) + \sigma_e^2/T_i][1 - \psi + \psi] = (1-\psi)^2\sigma_\mu^2 + \psi^2[\sigma_c^2(1-\gamma) + \sigma_e^2/T_i]$, which yields the final posterior:

$$f(\mu_i|\bar{g}_i, \bar{g}_{-i}) \propto exp\left\{-\frac{(\mu_i-\psi(\bar{g}_i-\gamma\bar{g}_{-i}))^2}{2[(1-\psi)^2\sigma_\mu^2+\psi^2[\sigma_c^2(1-\gamma)+\sigma_e^2/T_i]]}\right\} \quad \text{(A18)}$$

One will recognize Equation A18 as a proportional to a normal distribution with mean $\psi(\bar{g}_i - \gamma\bar{g}_{-i})$ and variance $RMSE^2 = (1-\psi)^2\sigma_\mu^2 + \psi^2[\sigma_c^2(1-\gamma) + \sigma_e^2/T_i]$. Hence $\mu_i|\bar{g}_i, \bar{g}_{-i} \sim N(\psi(\bar{g}_i - \gamma\bar{g}_{-i}), RMSE^2)$ as claimed. □

Part (2C) Credible interval

(C) From 2B, $\mu_i|\bar{g}_i, \bar{g}_{-i} \sim N(\hat{\mu}_i^{LS}, RMSE(\hat{\mu}_i)^2)$ and hence $(\mu_i - \hat{\mu}_i^{LS})/RMSE(\hat{\mu}_i) \sim N(0,1)$. From the standard normal:

$$P(-Z_{\alpha/2}^\star \le (\mu_i - \hat{\mu}_i)/RMSE(\hat{\mu}_i) \le Z_{\alpha/2}^\star) = 1-\alpha$$
$$P(-Z_{\alpha/2}^\star \times RMSE(\hat{\mu}_i) \le \mu_i - \hat{\mu}_i \le Z_{\alpha/2}^\star \times RMSE(\hat{\mu}_i)) = 1-\alpha$$
$$P(\hat{\mu}_i - Z_{\alpha/2}^\star \times RMSE(\hat{\mu}_i) \le \mu_i \le \hat{\mu}_i + Z_{\alpha/2}^\star \times RMSE(\hat{\mu}_i)) = 1-\alpha \quad □$$

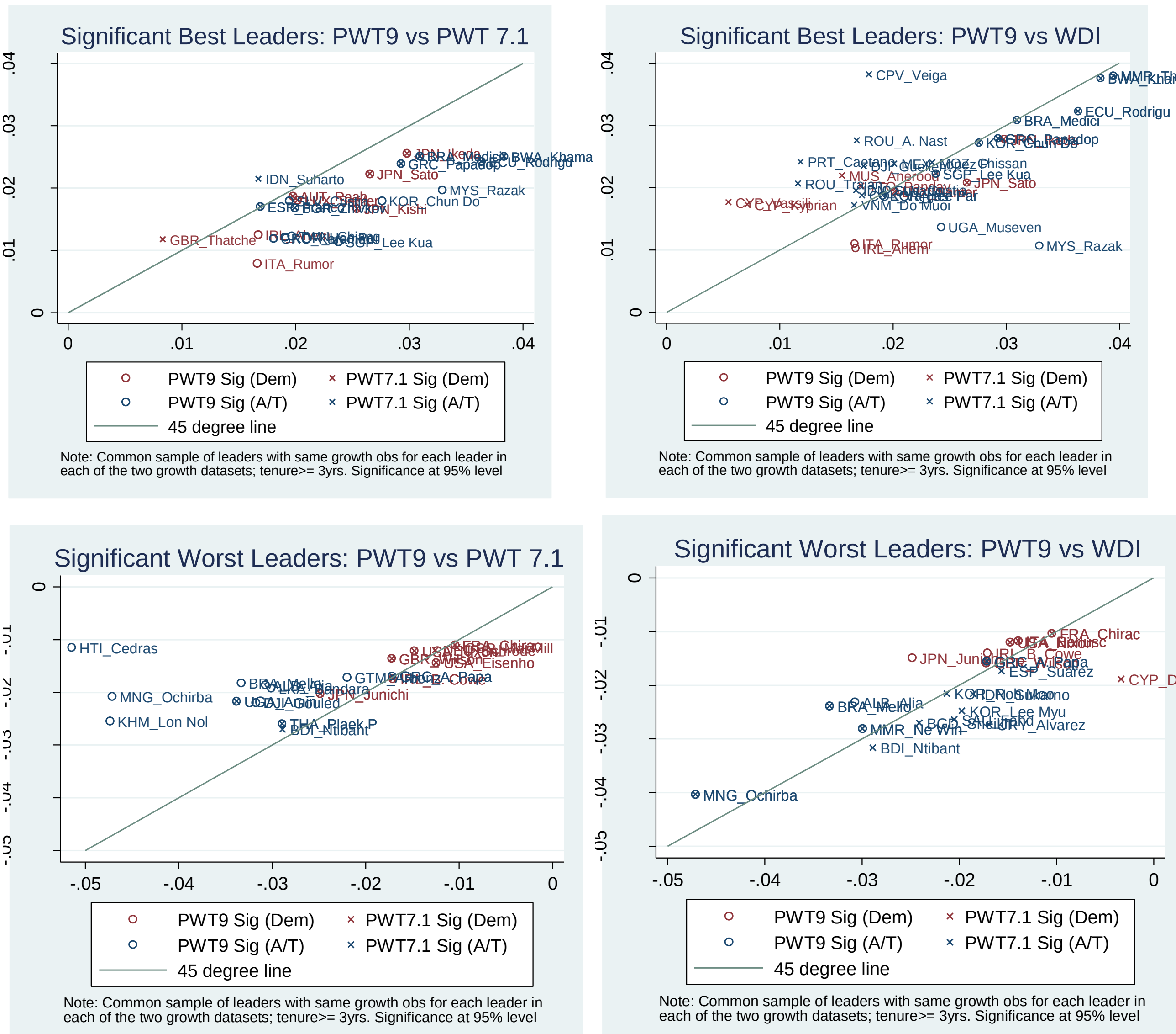
[47] That is, the distribution of μ_c after observing $\bar{g}_{-i}$ is $\mu_c|\bar{g}_{-i} \sim N(\gamma\bar{g}_{-i}, [\sigma_e^2/(N-T_i) + \sigma_\mu^2/L_{-i}]\gamma)$.

[48] The constant does not depend on μ_i because μ_i only shifts the mean (that is moves the distribution right or left) but does not affect the variance $\sigma_c^2(1-\gamma)(\sigma_e^2/T_i)/[\sigma_c^2(1-\gamma) + \sigma_e^2/T_i]$.

Appendix 2: Robustness to different growth datasets

Just as a teacher's value added can depend on the type of standardized test conducted, so leaders' least squares growth contribution can (and does) vary depending on the growth dataset used to measure it. In this section we compare the size and significance of least squares leader effects using two other datasets: PWT 7.1 and WDI (but applying the same least squares and SA methodology; only the underlying data varies). It turns out that the set of significant best and worst leaders changes substantially depending on the dataset used, which is one reason why we would not want to rely solely on growth data to judge the performance of leaders – even when using an optimal methodology. We prefer the results using the PWT9 as they improve upon PWT7.1 methodologically (as it is a later version), and has a longer and more complete sample than WDI.

Appendix Figure 1: Comparing the same set of best or worst leaders across growth datasets



There are three ways that different growth datasets can generate different leader effects. First, different datasets can generate different estimates of the underlying variation in leader quality σ_μ and other variance components (see Appendix Table 3), which feed into estimates of ψ and γ, which affect LS leader estimates. Specifically, the PWT9 data generate higher estimates of the underlying variation leader quality ($\sigma_\mu^{PWT9} = 1.7\%$) than either PWT 7.1

($\sigma_{\mu}^{PWT7.1} = 1.25\%$) or WDI ($\sigma_{\mu}^{WDI} = 1.5\%$). This is driven by more variation in leader quality in autocracies/transition countries, which leads to more autocrats being significantly good (or bad).[49]

Second, growth datasets can disagree on the raw leader growth average (or other leader growth average) for the same set of leaders and years. This is due to differences in methodologies and judgement calls when data quality is questionable.[50] This is a particularly big problem in lower income countries, which are often autocratic.

Finally, growth datasets cover different years and countries, resulting in missing leaders, or leaders with truncated tenures. Some of this is mechanical (and not problematic) because WDI growth data does not start until 1961 (rather than 1951 for PWT), and so is missing all potential best or worst leaders in the 1950s (such as Eisenhower). Likewise, PWT7.1 finishes in 2010, and so will truncate the tenures of more modern leaders. On the other hand, much of the data is inexplicably missing for whole countries (Afghanistan in PWT9 and Myanmar in PWT 7.1) or data starts late (WDI growth data for Haiti only starts in 1997, whereas PWT has data back 1961).[51]

The combined set of significantly best and worst leaders by different datasets is presented in Appendix Table 4 (best) and Appendix Table 5 (worst). While all three datasets agree on significantly good leaders like Khama, Rodriguez Lara, Medici, Ikeda, and Papadopoulos (significant and in the top 10 in all datasets), half of the top 10 leaders in PWT9 are either insignificant, lowly ranked or missing in at least one of the other two datasets.[52] Moreover, there is less agreement about the set of worst leaders: for bottom 10 significantly worst leaders in PWT9, none are significant in all three datasets (in part due to missing data), and only half are significant in one of the other two datasets. Most important here is the lower σ_{μ} and higher σ_{e}in PWT 7.1 (which reduces ψ and increases SEs), but also that there are many missing leaders -- especially in WDI. The later in part reflects the propensity for many extremely good or bad leaders to come from countries that experienced civil war, political turmoil or with questionable quality growth data which result in difficulties collecting reliable data on growth.

In Appendix Figure 1, we plot the size of significant leader effects for each data set, focusing on a set of leaders that have the full set of growth data (that is, we omit leaders where their tenure is truncated by the start or end of the sample for a particular dataset). While leader estimates are positively related, they are also a long way from 45 degree line, indicating substantial disagreement in the size of least squares leader effects. Comparing PWT9 vs PWT7.1, almost all the leader effects are closer to zero for PWT7.1 due to lower σ_{μ}, higher σ_{e} and hence more shrinkage (smaller ψ) in that dataset. This is particularly striking for the worst leader under PWT9 – Cedras (HTI) – which has leader effects 4 percentage points higher in PWT 7.1 (and insignificant), mostly explained by a raw leader growth average 6 percentage points higher under PWT7.1. PWT9 and WDI tend be in more agreement, with some of the best leaders (Khama, Than Shwe) and worst leaders (Ochibra) having similar sized leader effects.

[49] For autocracies/transitions countries $\sigma_{\mu}^{PWT9} = 2\%$ as against $\sigma_{\mu}^{PWT7.1} = 1.5\%$ and $\sigma_{\mu}^{WDI} = 1.7\%$. Estimates of σ_{e} (in general) and $\sigma_{\mu,DEM}$ are higher in PWT7.1 and lower in WDI (with much less disagreement about $\sigma_{\mu,DEM}$). The $\sigma_{c,AUT}$ is higher in WDI than elsewhere.

[50] For example, when Rwanda was recovering from genocide in 1995, log GDP per capita growth was around 30% in PWT9 and WDI, but around 60% (and hence an outlier) using PWT 7.1. As these data also depend on estimates of GDP per capita in 1994 during the genocide, they are highly questionable.

[51] Missing PWT 7.1 data for Myanmar meant Than Shwe was dropped from the dataset in the previous version of this paper.

[52] The disagreement among significant best leaders is not due to mechanical changes in tenure due to different start/end dates of different datasets. Even though the tenure of Hun Sen is shorter in WDI, this is due to missing data for Cambodia before 1994. For worst 10 leaders, only Bandaranaike, S.W.R.D. (Sri Lanka) and Plaek Pibulsongkram (Thailand) are mechanically missing due to a lack of WDI data in the 1950s – the other missing leaders in the bottom 10 are due to the late start of WDI growth data in Haiti, Cambodia, and Uganda, and missing growth data for Myanmar in PWT7.1.

However there are also major disagreements: Veiga (CPV) gets a leader effect of almost 4% in WDI, but 2% in PWT 9 (and is insignificant), largely because Veiga's leader growth average is 4 percentage points lower under PWT9. Similarly, Razak (MYS) gets a leader effect of 3.3% in PWT vs 1.1% in WDI, mostly explained by a growth average 3.5 percentage points lower under WDI.

Appendix 3: Additional Figures and Tables

Appendix Table 1A: Country List

Code	Name	Polity Ave	Established Democracy	Leaders	Obs*	Code	Name	Polity Ave	Established Democracy	Leaders	Obs*
ALB	Albania	-0.30	0	7	44	OMN	Oman	-9.24	0	1	44
DZA	Algeria	-5.13	0	6	53	PAK	Pakistan	1.19	0	15	64
AGO	Angola	-4.05	0	2	40	PAN	Panama	2.17	0	15	64
ARG	Argentina	1.34	0	18	64	PRY	Paraguay	-1.95	0	9	63
BHR	Bahrain	-8.93	0	2	44	PER	Peru	2.98	0	10	64
BGD	Bangladesh	0.77	0	8	44	PHL	Philippines	3.23	0	10	64
BEN	Benin	-0.20	0	9	55	POL	Poland	2.50	0	7	44
BTN	Bhutan	-7.50	0	8	44	PRT	Portugal	2.77	0	8	64
BOL	Bolivia (Plurina	2.05	0	18	64	QAT	Qatar	-10.00	0	4	44
BWA	Botswana	6.94	0	4	49	KOR	Republic of Korea	1.49	0	11	61
BRA	Brazil	2.78	0	16	64	ROU	Romania	-0.31	0	10	54
BGR	Bulgaria	1.84	0	11	44	RWA	Rwanda	-5.44	0	3	54
BFA	Burkina Faso	-3.78	0	7	55	SAU	Saudi Arabia	-10.00	0	4	44
BDI	Burundi	-2.62	0	7	53	SEN	Senegal	-0.35	0	4	54
CPV	Cabo Verde	4.08	0	3	40	SLE	Sierra Leone	-0.89	0	9	53
KHM	Cambodia	-0.23	0	5	44	SGP	Singapore	-1.65	0	3	54
CMR	Cameroon	-5.98	0	2	54	ZAF	South Africa	5.77	0	9	64
CAF	Central African	-3.37	0	7	54	ESP	Spain	3.14	0	7	64
TCD	Chad	-4.72	0	5	54	LKA	Sri Lanka	5.95	0	10	64
CHL	Chile	3.75	0	11	63	SDN	Sudan (Former)	-3.98	0	4	44
CHN	China	-7.42	0	6	62	SUR	Suriname	2.80	0	6	40
COL	Colombia	6.27	0	17	64	SWZ	Swaziland	-9.05	0	4	44
COM	Comoros	1.35	0	7	40	SYR	Syrian Arab Republic	-8.04	0	6	54
COG	Congo	-4.76	0	6	54	TWN	Taiwan	-1.02	0	6	63
CIV	Cote d'Ivoire	-5.48	0	5	54	THA	Thailand	0.95	0	20	64
COD	D.R. of the Con	-3.58	0	4	55	TGO	Togo	-4.93	0	4	54
DJI	Djibouti	-3.42	0	2	38	TUN	Tunisia	-5.52	0	4	54
DOM	Dominican Rep	2.37	0	11	63	TUR	Turkey	6.53	0	16	64
ECU	Ecuador	3.98	0	20	64	TZA	U.R. of Tanzania: Mai	-4.09	0	4	54
EGY	Egypt	-5.81	0	8	64	UGA	Uganda	-2.81	0	5	53
SLV	El Salvador	2.14	0	16	64	ARE	United Arab Emirates	-8.00	0	2	44
GNQ	Equatorial Guir	-6.28	0	3	52	URY	Uruguay	5.61	0	23	64
ETH	Ethiopia	-7.58	0	5	64	VEN	Venezuela (Bolivarian	5.69	0	10	64
FJI	Fiji	4.40	0	5	45	VNM	Viet Nam	-7.00	0	6	44
GAB	Gabon	-5.69	0	3	54	ZMB	Zambia	-0.86	0	5	51
GMB	Gambia	2.26	0	2	50	ZWE	Zimbabwe	-0.93	0	3	50
GHA	Ghana	-1.11	0	11	59	AUS	Australia	10.00	1	12	64
GRC	Greece	6.19	0	15	63	AUT	Austria	10.00	1	11	64
GTM	Guatemala	1.36	0	18	64	BEL	Belgium	9.75	1	15	64
GIN	Guinea	-5.25	0	5	55	CAN	Canada	10.00	1	9	64
GNB	Guinea-Bissau	-1.41	0	8	41	CRI	Costa Rica	10.00	1	15	64
HTI	Haiti	-4.11	0	11	54	CYP	Cyprus	8.65	1	7	55
HND	Honduras	2.77	0	18	64	DNK	Denmark	10.00	1	13	64
HUN	Hungary	3.02	0	9	44	FIN	Finland	10.00	1	6	64
IDN	Indonesia	-2.48	0	6	54	FRA	France	8.16	1	13	64
IRN	Iran (Islamic Re	-6.39	0	6	59	DEU	Germany	10.00	1	8	64
IRQ	Iraq	-6.87	0	6	44	IND	India	8.66	1	12	64
JOR	Jordan	-6.10	0	2	60	IRL	Ireland	9.97	1	12	64
KEN	Kenya	-1.58	0	4	52	ISR	Israel	7.66	1	12	64
KWT	Kuwait	-8.02	0	3	42	ITA	Italy	10.00	1	24	64
LAO	Lao People's DI	-6.52	0	5	44	JAM	Jamaica	9.58	1	7	53
LBN	Lebanon	2.76	0	8	44	JPN	Japan	10.00	1	25	64
LSO	Lesotho	0.51	0	6	49	LUX	Luxembourg	10.00	1	8	64
LBR	Liberia	-2.13	0	10	50	MUS	Mauritius	9.70	1	4	47
MDG	Madagascar	0.57	0	7	54	NLD	Netherlands	10.00	1	12	64
MWI	Malawi	-2.90	0	5	51	NZL	New Zealand	10.00	1	13	64
MYS	Malaysia	5.14	0	6	58	NOR	Norway	10.00	1	13	64
MLI	Mali	-1.20	0	6	54	SWE	Sweden	10.00	1	8	64
MRT	Mauritania	-5.78	0	7	54	CHE	Switzerland	10.00	1	45	64
MEX	Mexico	-0.61	0	12	64	TTO	Trinidad and Tobago	8.92	1	6	53
MNG	Mongolia	2.20	0	6	44	GBR	United Kingdom	10.00	1	13	64
MAR	Morocco	-6.83	0	3	59	USA	United States	10.00	1	12	64
MOZ	Mozambique	-0.90	0	3	40	NOR	Norway	10.00	1	13	64
MMR	Myanmar	-6.43	0	4	44	SWE	Sweden	10.00	1	8	64
NPL	Nepal	-1.96	0	15	54	CHE	Switzerland	10.00	1	45	64
NIC	Nicaragua	-0.73	0	8	64	TTO	Trinidad and Tobago	8.92	1	6	53
NER	Niger	-2.15	0	9	54	GBR	United Kingdom	10.00	1	13	64
NGA	Nigeria	-0.49	0	11	55	USA	United States	10.00	1	12	64

Notes: *Number of observations with PWT9 growth data. "Established Democracies" have Average Polity Score>7.5

Appendix Table 1B: Outliers based PWT9 Growth Rates					
			Growth rate	Comparison growth	
Country Name	Country Code	Year	PWT 9	PWT 7.1	WDI
Algeria	DZA	1962	-45.9%	-45.2%	-24.4%
Central African Republic	CAF	2013	-46.7%	Missing	-47.7%
Equatorial Guinea	GNQ	1997	63.5%	76.7%	88.2%
Equatorial Guinea	GNQ	2001	48.4%	38.6%	45.7%
Iran (Islamic Republic of)	IRN	1970	-52.6%	14.0%	7.7%
Iraq	IRQ	1991	-110.9%	-103.9%	-105.0%
Iraq	IRQ	2003	-42.9%	-44.4%	-42.9%
Iraq	IRQ	2004	40.6%	41.9%	40.6%
Kuwait	KWT	1990	Dropped because of Iraqi invasion of		
		1991		Kuwait	
		1992	65.3%	13.5%	Missing
Liberia	LBR	All	Influential country*		
Lebanon	LBN	1976	-85.3%	-81.0%	Missing
Lebanon	LBN	1977	60.3%	55.7%	Missing
Lebanon	LBN	1982	-46.5%	-58.1%	Missing
Lebanon	LBN	1989	-55.6%	-57.7%	-55.6%
Mauritania	MRT	1964	41.8%	42.6%	21.5%
Myanmar	MMR	<=1970	GDPPC in Tripled 1970		
Rwanda	RWA	1994	-62.0%	-70.9%	-64.9%
Zimbabwe	ZWE	2009	42.5%	8.6%	4.2%

Notes: Outliers are dropped when the abs(log-per capita growth)>40% in a given year. Listed data are outliers based on PWT9 growth rates (outliers for PWT 7.1 and WDI generated in an analagous way). *Liberia is dropped due to many influential observations during the first and second Liberian Civil wars.

Appendix Table 2: Monte Carlo Estimates - alternate methods/samples (when σs are unknown)							
	True: sd(leader)=1.5%; sd(country)=1%; sd(iid)=5% (approximately)						
	Estimated Variance components			Shrinkage Coeffs		Performance	
	SD(leader)	sd(iid)	sd(CE)	ψ (shrinkage)	γ (country adj)	Bias (λ≠1)	Std Error
	Panel A: SA Method, normal errors (homoskasticity)						
(A1) Simple model SA method	1.50%	5.00%					
(country sd=0, 500 reps)	[0.09%]	[0.04%]	-		-		
(A2) SA method full model	1.77%	5.00%	0.99%	0.34	0.53	0.856033	1.28%
(country sd=0, 500 reps)	[0.11%]	[0.04%]	[0.06%]	(0.17)	(0.1)	[0.08]	[0.03%]
	Panel B: RE Method, normal errors (homoskasticity)						
(B) RE method full model (500	1.55%	5.00%	1.40%	0.28	0.71	1.01	1.29%
reps)	[0.22%]	[0.05%]	[0.08%]	(0.14)	(0.09)	[0.23]	[0.03%]
	Panel C: SA Method, leptokurtic errors (homoskasticity)						
C) SA method full model (500	1.77%	5.00%	1.40%	0.33	0.69	0.84	1.30%
reps)	[0.14%]	[0.09%]	[0.09%]	(0.16)	(0.1)	[0.1]	[0.03%]
	Panel D: Leptokurtic errors and heteroskasticity by aut/dem						
D1) SA method full model (500	1.67%	4.90%	1.40%	0.31	0.70	0.93	1.27%
reps)	[0.14%]	[0.09%]	[0.09%]	(0.15)	(0.09)	[0.12]	[0.03%]
D2) RE method full model (500	1.15%	4.90%	1.41%	0.20	0.74	2.49	1.32%
reps)	[0.48%]	[0.09%]	[0.09%]	(0.11)	(0.08)	[10.87]	[0.07%]
	Panel E: SA Method, leptokurtic errors (aut/dem separately)						
E1) Autocracies only SA method	1.84%	5.30%	1.46%	0.33	0.66	0.81	1.31%
full model (500 reps)	[0.17%]	[0.11%]	[0.1%]	(0.16)	(0.1)	[0.12]	[0.04%]
E2) Democracies only SA method	1.56%	2.89%	1.17%	0.46	0.79	0.96	1.13%
full model (500 reps)	[0.13%]	[0.1%]	[0.16%]	(0.17)	(0.06)	[0.11]	[0.05%]

* Note: Table presents Monte Carlo estimates of leader effects and variance components, where the actual country X leader tenure structure is used, but the growth process is drawn from a normal distribution. A successful method uncovers the "true" parameter of the leader effect of 1.5%, CE of 1% and iid SE of 5%, is unbiased (λ=1) and has a lowest standard error. Calculations of variance components use the "SA" method a described in the text. Standard deviations across Monte Carlo replications are reported in brackets and standard deviations across leaders are reported in parentheses.

Appendix Table 3: Estimates of Variance Components (alternative methods and datasets)

		Pooled			Autocracies/Transition			Established Democracies		
Dataset	**Method**	**sigmaU**	**sigmaC**	**sigmaE***	**sigmaU**	**sigmaC**	**sigmaE***	**sigmaU**	**sigmaC**	**sigmaE***
PWT9	SA^	1.69%	1.27%	4.60%	1.96%	1.39%	5.01%	1.11%	0.72%	2.65%
	RE	0.88%	1.23%	4.60%	1.45%	1.36%	5.01%	0.82%	0.72%	2.65%
PWT7.1	SA	1.23%	1.26%	5.40%	1.49%	1.36%	5.85%	1.03%	0.72%	3.25%
	RE	0.00%	1.27%	5.40%	0.00%	1.37%	5.85%	0.46%	0.75%	3.25%
WDI	SA	1.53%	1.35%	4.29%	1.72%	1.47%	4.68%	1.21%	0.71%	2.20%
	RE	0.90%	1.34%	4.29%	1.17%	1.45%	4.68%	1.23%	0.71%	2.20%

Notes: ^Repeated from main text. * SigmaE is the across-sample SD of the iid error. In the estimation of leader effects, we allow sigmaE to vary across countries. Method: SA is random effects adjusting for unbalanced panels (default method from the main text). RE is standard random effects without unabalanced panel adjustment. Outliers are dropped in all specification.

Appendix Table 4: Significantly Best Leaders at the 95% level in any of the three datasets.

Name	**Country**	**Dem**	**Rank**			**Significant 95%**			**LS Leader Estimate**			**Error**		
			PWT9	PWT71	WDI	PWT9	PWT71	WDI	PWT9	PWT71	WDI	PWT9	PWT71	WDI
Than Shwe	MMR	0	1		2	1		1	3.9%		3.8%	1.1%		1.0%
Khama	BWA	0	2	4	3	1	1	1	3.8%	2.5%	3.8%	1.3%	1.2%	1.1%
Rodriguez Lara	ECU	0	3	5	4	1	1	1	3.6%	2.4%	3.2%	1.3%	1.1%	1.0%
Razak	MYS	0	4	9	89	1	0	0	3.3%	2.0%	1.1%	1.3%	1.2%	1.1%
Medici	BRA	0	5	3	5	1	1	1	3.1%	2.5%	3.1%	1.1%	1.1%	1.1%
Ikeda	JPN	1	6	1	8	1	1	1	3.0%	2.6%	2.8%	0.8%	0.8%	0.9%
Papadopoulos	GRC	0	7	6	7	1	1	1	2.9%	2.4%	2.8%	1.0%	0.9%	0.9%
Hun Sen	KHM	0	8	2	59	1	1	0	2.9%	2.5%	1.5%	1.1%	1.1%	1.0%
Chun Doo Hwan	KOR	0	9	15	10	1	0	1	2.8%	1.8%	2.7%	1.0%	1.0%	0.9%
Sato	JPN	1	10	7	19	1	1	1	2.7%	2.2%	2.1%	0.6%	0.6%	0.8%
Kishi	JPN	1	12	22		1	1		2.5%	1.7%		0.8%	0.8%	
Faisal	SAU	0	15		6	0		1	2.4%		2.9%	1.8%		1.4%
Museveni	UGA	0	16	11	61	1	1	0	2.4%	1.9%	1.4%	0.9%	0.9%	1.1%
Lee Kuan Yew	SGP	0	17	58	17	1	0	1	2.4%	1.1%	2.2%	1.0%	0.9%	1.0%
Chissano	MOZ	0	18	10	12	0	0	1	2.3%	2.0%	2.4%	1.2%	1.0%	1.2%
Adenauer	DEU	1	21			1			2.2%			0.6%		
Santer	LUX	1	27	14	29	1	1	1	2.0%	1.8%	1.9%	0.7%	0.8%	0.8%
Lopez Portillo	MEX	0	28	76	13	0	0	1	2.0%	1.0%	2.4%	1.1%	1.1%	1.0%
Zhivkov	BGR	0	30	21	137	1	1	0	2.0%	1.7%	0.8%	0.9%	0.8%	1.0%
Raab	AUT	1	32	12		1	1		2.0%	1.9%		0.6%	0.6%	
Chiang Ching-Kuo	TWN	0	33	45		1	0		2.0%	1.2%		0.9%	0.9%	
Cristiani	SLV	0	34	16	27	1	0	0	1.9%	1.8%	2.0%	1.0%	0.9%	1.1%
Hee Park	KOR	0	38	48	33	1	0	1	1.9%	1.2%	1.9%	0.8%	0.8%	0.8%
Karamanlis	GRC	0	40	51	34	1	0	1	1.8%	1.2%	1.8%	0.8%	0.8%	0.8%
Veiga	CPV	0	41	60	1	0	0	1	1.8%	1.1%	3.8%	1.3%	1.2%	1.1%
Guelleh	DJI	0	42	160	14	0	0	1	1.7%	0.6%	2.4%	1.3%	1.2%	1.0%
Vargas	COL	0	45	25	32	0	0	1	1.7%	1.6%	1.9%	0.9%	0.9%	0.8%
Panday	TTO	1	47	238	23	0	0	1	1.7%	0.4%	2.0%	0.9%	1.0%	1.0%
Franco	ESP	0	51	20	36	1	1	1	1.7%	1.7%	1.8%	0.9%	0.8%	0.7%
A. Nastase	ROU	0	52	64	9	0	0	1	1.7%	1.1%	2.8%	1.4%	1.3%	1.1%
Suharto	IDN	0	53	8	28	0	1	1	1.7%	2.1%	2.0%	0.9%	0.8%	0.8%
Ahern	IRL	1	54	44	95	1	0	0	1.7%	1.3%	1.0%	0.6%	0.7%	0.7%
Rumor	ITA	1	56	105	83	1	0	0	1.7%	0.8%	1.1%	0.7%	0.7%	0.7%
Do Muoi	VNM	0	57	40	40	0	0	1	1.7%	1.3%	1.7%	1.1%	1.0%	0.8%
Anerood Jugnauth	MUS	1	66	69	18	0	0	1	1.5%	1.0%	2.2%	0.8%	0.9%	0.7%
Caetano	PRT	0	104	26	11	0	0	1	1.2%	1.5%	2.4%	1.0%	1.0%	1.0%
Traian Basescu	ROU	0	106	227	20	0	0	1	1.2%	0.4%	2.1%	1.1%	1.2%	1.0%
Ayub Khan	PAK	0	120	151	24	0	0	1	1.1%	0.6%	2.0%	0.8%	0.9%	0.8%
Verwoerd	ZAF	0	150	262	31	0	0	1	0.9%	0.3%	1.9%	0.9%	0.9%	0.9%
Thatcher	GBR	1	159	53	203	0	1	0	0.8%	1.2%	0.5%	0.5%	0.5%	0.5%
Kyprianou	CYP	1	187	197	39	0	0	1	0.7%	0.5%	1.7%	1.0%	1.0%	0.6%
Vassiliou	CYP	1	234	245	37	0	0	1	0.5%	0.3%	1.8%	1.0%	1.0%	0.7%

Note: List of leaders with tenure>=3yrs which are statistically significant in any of the 3 datasets. Blank indicates leader is missing in dataset. Leaders are ordered by their PWT9 leader estimate as in the main text.

Appendix Table 4B: Significantly Best Leaders at the 90% level (tenure 3+ yrs)											
Rank	**Name**	**Country**	**LS Leader Est.**	**RMS Error**	**Raw Growth Ave**	**Shrinkage (ψ)**	**Sig95%**	**Sig 99%**	**Tenure**	**Dem**	**1stYear**
1	Than Shwe	MMR	3.95%	1.06%	8.58%	0.71	1	1	19	0	1992
2	Khama	BWA	3.83%	1.30%	9.56%	0.56	1	1	15	0	1966
3	Rodriguez Lara	ECU	3.63%	1.29%	9.14%	0.57	1	1	4	0	1972
4	Razak	MYS	3.29%	1.28%	7.47%	0.58	1	1	5	0	1971
5	Medici	BRA	3.09%	1.12%	8.71%	0.67	1	1	4	0	1970
6	Ikeda	JPN	2.98%	0.77%	8.65%	0.52	1	1	4	1	1961
7	Papadopoulos	GRC	2.93%	0.99%	8.21%	0.75	1	1	6	0	1968
8	Hun Sen	KHM	2.90%	1.13%	5.07%	0.67	1	0	26	0	1985
9	Chun Doo Hwan	KOR	2.76%	1.03%	8.09%	0.72	1	1	7	0	1981
10	Sato	JPN	2.65%	0.65%	7.70%	0.66	1	1	8	1	1965
11	Nkurunziza	BDI	2.63%	1.38%	7.45%	0.50	0	0	9	0	2006
12	Kishi	JPN	2.54%	0.77%	7.81%	0.52	1	1	4	1	1957
13	Quirrino	PHL	2.45%	1.32%	5.25%	0.55	0	0	3	0	1951
14	Museveni	UGA	2.42%	0.92%	3.25%	0.78	1	1	29	0	1986
15	Lee Kuan Yew	SGP	2.38%	1.03%	6.26%	0.73	1	0	30	0	1961
16	Chissano	MOZ	2.34%	1.21%	4.26%	0.62	0	0	18	0	1987
17	Paul Kagame	RWA	2.33%	1.31%	6.08%	0.55	0	0	20	0	1995
18	Masire	BWA	2.29%	1.27%	5.87%	0.58	0	0	17	0	1981
19	Adenauer	DEU	2.22%	0.57%	6.34%	0.73	1	1	13	1	1951
20	Saksgoburggotski	BGR	2.18%	1.29%	6.84%	0.57	0	0	4	0	2002
21	Yen Chia-Kan	TWN	2.09%	1.21%	7.47%	0.62	0	0	3	0	1975
22	Santer	LUX	2.02%	0.75%	4.72%	0.55	1	1	10	1	1985
23	Lopez Portillo	MEX	2.01%	1.11%	3.51%	0.68	0	0	6	0	1977
24	Zhivkov	BGR	1.99%	0.91%	5.08%	0.78	1	0	19	0	1971
25	Kubitschek	BRA	1.99%	1.05%	4.79%	0.71	0	0	5	0	1956
26	Raab	AUT	1.98%	0.58%	5.89%	0.72	1	1	8	1	1953
27	Chiang Ching-Kuo	TWN	1.97%	0.94%	6.82%	0.77	1	0	10	0	1978
28	Cristiani	SLV	1.94%	0.98%	3.15%	0.75	1	0	5	0	1989
29	Roh Tae Woo	KOR	1.94%	1.13%	7.65%	0.67	0	0	5	0	1988
30	Hee Park	KOR	1.91%	0.82%	6.42%	0.82	1	0	19	0	1961
31	Karamanlis	GRC	1.81%	0.81%	4.87%	0.83	1	0	13	0	1956
32	Mendez Montenegr	GTM	1.73%	0.91%	3.56%	0.78	0	0	4	0	1966
33	Rivera	SLV	1.73%	0.98%	4.22%	0.75	0	0	5	0	1962
34	Vargas	COL	1.72%	0.88%	2.22%	0.80	0	0	4	0	1987
35	Panday	TTO	1.72%	0.91%	6.52%	0.33	0	0	6	1	1996
36	Betancur	COL	1.72%	0.88%	1.22%	0.80	0	0	4	0	1983
37	Franco	ESP	1.69%	0.86%	5.42%	0.81	1	0	25	0	1951
38	Suharto	IDN	1.67%	0.89%	4.36%	0.80	0	0	32	0	1966
39	Ahern	IRL	1.67%	0.62%	5.32%	0.69	1	1	11	1	1997
40	Rumor	ITA	1.66%	0.69%	6.85%	0.61	1	0	4	1	1969
41	Yoshida, Shigeru	JPN	1.58%	0.82%	6.29%	0.45	0	0	3	1	1952
42	Anerood Jugnauth	MUS	1.55%	0.85%	4.22%	0.41	0	0	17	1	1982
43	Fanfani	ITA	1.25%	0.65%	4.96%	0.66	0	0	5	1	1958
44	de Gasperi	ITA	1.24%	0.75%	5.60%	0.55	0	0	3	1	1951
45	Haughey	IRL	1.20%	0.69%	3.30%	0.61	0	0	7	1	1980
46	Thatcher	GBR	0.83%	0.49%	2.40%	0.80	0	0	12	1	1979

Notes: list of leaders (tenure >=3yrs) which have significantly positive LS leader effect at the 10% level (two-tailed). That is where (LS leader estimate)-Z*(RMS Err)>0 for Z=1.65. Sig 99% is analagous, but where Z=2.576.

Appendix Table 5: Significantly Worst Leaders at the 95% level in any of the three datasets.

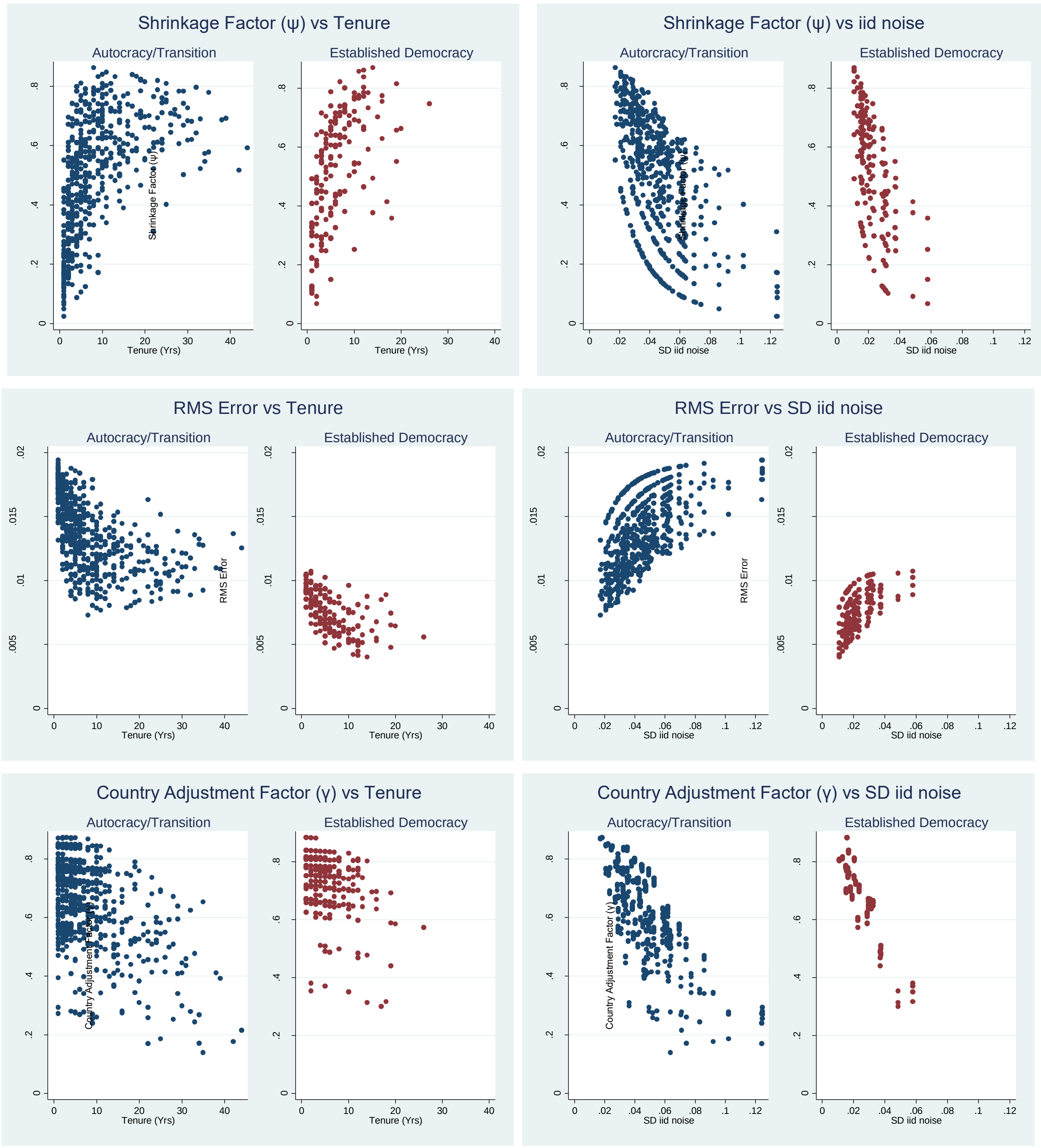
Name	Country	Dem	Rank (from worst)			Significant 95%			LS Leader Estimate			Error		
			PWT9	PWT71	WDI	PWT9	PWT71	WDI	PWT9	PWT71	WDI	PWT9	PWT71	WDI
Cedras	HTI	0	1	58		-1	0		-5.1%	-1.1%		1.4%	1.3%	
Lon Nol	KHM	0	2	3		-1	0		-4.7%	-2.5%		1.6%	1.3%	
Ochirbat	MNG	0	3	7	2	-1	0	-1	-4.7%	-2.1%	-4.0%	1.2%	1.3%	1.2%
Amin	UGA	0	4	6		-1	-1		-3.4%	-2.2%		1.1%	1.1%	
Mello	BRA	0	5	16	10	-1	0	-1	-3.3%	-1.8%	-2.4%	1.2%	1.1%	1.2%
Gouled Aptidon	DJI	0	7	5	1	-1	0	-1	-3.2%	-2.2%	-4.4%	1.3%	1.2%	1.1%
Alia	ALB	0	8	11	11	-1	0	0	-3.1%	-1.9%	-2.3%	1.6%	1.3%	1.5%
Bandaranaike, S.W.R.D.	LKA	0	9	9		-1	0		-3.0%	-1.9%		1.3%	1.2%	
Ne Win	MMR	0	10		4	-1		-1	-3.0%		-2.8%	1.1%		1.0%
Plaek Pibulsongkram	THA	0	11	2		-1	-1		-2.9%	-2.6%		1.4%	1.1%	
Ntibantunganya	BDI	0	12	1	3	0	-1	-1	-2.9%	-2.7%	-3.2%	1.7%	1.3%	1.4%
Khalifah Ath-Thani	QAT	0	13	140		-1	0		-2.7%	-0.6%		1.2%	1.3%	
Junichiro Koizumi	JPN	1	16	8	45	-1	-1	0	-2.5%	-2.0%	-1.5%	0.7%	0.7%	0.9%
Sheikh Mujib Rahman	BGD	0	17	4	6	0	0	-1	-2.4%	-2.2%	-2.7%	1.3%	1.1%	1.2%
Arbenz Guzman	GTM	0	22	20		-1	0		-2.2%	-1.7%		0.9%	1.0%	
Ma Ying-jeou	TWN	0	23	34		-1	0		-2.2%	-1.4%		1.0%	1.1%	
Roh Moo Hyun	KOR	0	26	38	14	0	0	-1	-2.1%	-1.3%	-2.2%	1.1%	1.1%	1.0%
Fahd	SAU	0	28	356	8	0	0	-1	-2.1%	0.0%	-2.6%	1.5%	1.2%	1.2%
Lee Myung Bak	KOR	0	30	62	9	0	0	-1	-2.0%	-1.1%	-2.5%	1.1%	1.2%	1.0%
Sukarno	IDN	0	33	10	13	0	0	-1	-1.9%	-1.9%	-2.2%	1.2%	1.1%	1.0%
Wilson	GBR	1	39	36	37	-1	-1	-1	-1.7%	-1.4%	-1.6%	0.6%	0.5%	0.5%
A. Papandreou	GRC	0	40	22	40	-1	-1	-1	-1.7%	-1.7%	-1.6%	0.9%	0.8%	0.8%
B. Cowen	IRL	1	41	19	55	-1	-1	0	-1.7%	-1.7%	-1.4%	0.8%	0.9%	1.0%
Alvarez Armalino	URY	0	44	12	5	0	0	-1	-1.7%	-1.9%	-2.7%	1.4%	1.3%	1.2%
Suarez Gonzalez	ESP	0	50	25	29	0	0	-1	-1.6%	-1.5%	-1.7%	1.2%	1.0%	0.9%
Nixon	USA	1	59	49	84	-1	-1	-1	-1.5%	-1.2%	-1.2%	0.6%	0.6%	0.6%
Berlusconi	ITA	1	61	26	86	-1	-1	-1	-1.4%	-1.5%	-1.2%	0.5%	0.5%	0.5%
Schroder	DEU	1	78	52	69	0	-1	0	-1.3%	-1.2%	-1.3%	0.7%	0.6%	0.7%
Eisenhower	USA	1	80	31		-1	-1		-1.2%	-1.4%		0.6%	0.6%	
Chirac	FRA	1	109	61	99	-1	-1	-1	-1.0%	-1.1%	-1.0%	0.4%	0.4%	0.5%
MacMillan	GBR	1	111	59	72	0	-1	0	-1.0%	-1.1%	-1.2%	0.6%	0.5%	0.7%
Dimitris Christofias	CYP	1	251	331	23	0	0	-1	-0.3%	0.0%	-1.9%	1.0%	1.0%	0.7%
Zardari	PAK	0	741		12			-1			-2.2%			0.9%

Note: List of leaders with tenure>=3yrs which are statistically significant in any of the 3 datasets. Blank indicates leader is missing in dataset. Leaders are ordered by their PWT9 leader estimate as in the main text.

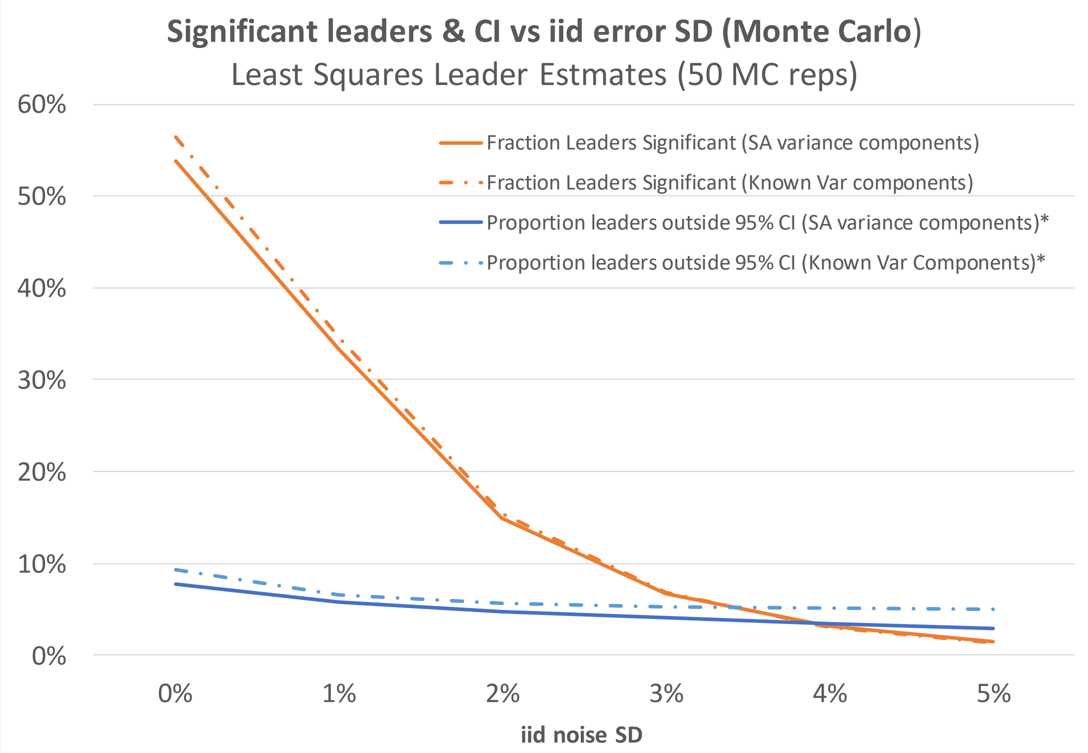
Appendix Table 5B: Significantly Worst Leaders at the 90% level (tenure 3+ yrs)											
Rank	**Name**	**Country**	**LS Leader Est.**	**RMS Error**	**Raw Growth Ave**	**Shrinkage (ψ)**	**Sig 95%**	**Sig 99%**	**Tenure**	**Dem**	**1stYear**
1	Cedras	HTI	-5.15%	1.39%	-9.85%	0.50	-1	-1	3	0	1992
2	Lon Nol	KHM	-4.74%	1.56%	-11.12%	0.37	-1	-1	4	0	1971
3	Ochirbat	MNG	-4.71%	1.20%	-3.21%	0.62	-1	-1	7	0	1990
4	Antall	HUN	-3.58%	1.05%	-4.87%	0.72	-1	-1	4	0	1990
5	Amin	UGA	-3.38%	1.10%	-2.97%	0.69	-1	-1	8	0	1971
6	Mello	BRA	-3.33%	1.21%	-2.97%	0.62	-1	-1	3	0	1990
7	Oueddei	TCD	-3.30%	1.74%	-15.98%	0.21	0	0	3	0	1979
8	Gouled Aptidon	DJI	-3.18%	1.25%	-4.28%	0.59	-1	0	22	0	1977
9	Alia	ALB	-3.07%	1.56%	-5.81%	0.37	-1	0	7	0	1985
10	Bandaranaike, S.W.F	LKA	-3.01%	1.28%	-2.85%	0.57	-1	0	4	0	1956
11	Ne Win	MMR	-3.00%	1.07%	0.36%	0.70	-1	-1	18	0	1971
12	Plaek Pibulsongkram	THA	-2.89%	1.40%	-2.89%	0.49	-1	0	7	0	1951
13	Ntibantunganya	BDI	-2.89%	1.66%	-8.16%	0.28	0	0	3	0	1994
14	Khalifah Ath-Thani	QAT	-2.74%	1.22%	-3.19%	0.61	-1	0	23	0	1972
15	Yameogo	BFA	-2.67%	1.39%	-2.50%	0.50	0	0	6	0	1960
16	Mohammed V	MAR	-2.58%	1.49%	-0.08%	0.42	0	0	5	0	1956
17	Junichiro Koizumi	JPN	-2.48%	0.70%	1.08%	0.61	-1	-1	6	1	2001
18	Sheikh Mujib Rahma	BGD	-2.41%	1.32%	-3.45%	0.55	0	0	4	0	1972
19	Anastasio Somoza D	NIC	-2.40%	1.34%	-2.06%	0.53	0	0	13	0	1967
20	Arbenz Guzman	GTM	-2.20%	0.91%	-0.99%	0.78	-1	0	4	0	1951
21	Ma Ying-jeou	TWN	-2.17%	1.00%	2.69%	0.74	-1	0	7	0	2008
22	Antonis Konstantino	GRC	-2.16%	1.20%	-2.93%	0.63	0	0	3	0	2012
23	Roh Moo Hyun	KOR	-2.13%	1.13%	3.77%	0.67	0	0	5	0	2003
24	Machel	MOZ	-2.12%	1.28%	-2.65%	0.57	0	0	12	0	1975
25	Lee Myung Bak	KOR	-1.97%	1.13%	2.56%	0.67	0	0	5	0	2008
26	Mobutu	COD	-1.90%	1.11%	-3.20%	0.68	0	0	31	0	1966
27	Zardari	PAK	-1.85%	1.01%	0.54%	0.74	0	0	4	0	2009
28	Shinzo Abe	JPN	-1.83%	0.82%	1.28%	0.45	-1	0	3	1	2007
29	Wilson	GBR	-1.72%	0.57%	0.78%	0.74	-1	-1	7	1	1965
30	A. Papandreou	GRC	-1.72%	0.86%	0.68%	0.81	-1	0	10	0	1982
31	B. Cowen	IRL	-1.71%	0.84%	-4.22%	0.43	-1	0	3	1	2008
32	Mahendra	NPL	-1.56%	0.94%	0.41%	0.77	0	0	11	0	1961
33	Nixon	USA	-1.48%	0.61%	1.83%	0.70	-1	0	6	1	1969
34	Berlusconi	ITA	-1.39%	0.53%	-0.12%	0.77	-1	-1	10	1	1994
35	de Valera	IRL	-1.32%	0.75%	0.82%	0.54	0	0	5	1	1951
36	Schroder	DEU	-1.26%	0.67%	1.24%	0.64	0	0	7	1	1999
37	Eisenhower	USA	-1.25%	0.56%	1.23%	0.74	-1	0	8	1	1953
38	Diefenbaker	CAN	-1.20%	0.67%	1.05%	0.64	0	0	6	1	1957
39	Menzies	AUS	-1.05%	0.61%	1.82%	0.70	0	0	15	1	1951
40	Chirac	FRA	-1.05%	0.41%	1.72%	0.86	-1	0	12	1	1995
41	MacMillan	GBR	-1.02%	0.57%	1.68%	0.74	0	0	7	1	1957
42	Gerhardsen	NOR	-0.87%	0.51%	2.72%	0.79	0	0	12	1	1951

Notes: list of leaders (tenure 3+ yrs) who have sign negative LS leader effect at the 10% level (two-tailed). That is if (LS leader est.)+Z*(RMS Err)<0 for Z=1.645. Sig 99% is analagous, but where Z=2.576. Rank from worst.

Appendix Figure 2



Appendix Figure 3



Note: * a 95% CI should have 5% of leaders outside the CI.

Appendix 4: Data Sources

Leader Data

Leader data comes from Archigos 4.1 (Goemans et al 2009), downloaded from http://privatewww.essex.ac.uk/~ksg/archigos.html (data file: arch_annual.txt, accessed 15 March 2017). In the case there are multiple leaders in office in a a particular year, the year is allocated to the leader in office for the most number of days.[53]

Polity IV Data (Democracy vs Autocracy)

Polity IV data comes from: http://www.systemicpeace.org/inscr/p4v2015.xls (accessed 15 March 2017). We calculated the average Polity score over our sample, with established democracies having an average polity score >=7.5, and autocracies/transition countries have an average polity score <7.5. Countries with no Polity data for the whole sample were dropped.

PWT Growth Data

We use two versions of PWT data: PWT 7.1, and PWT9. PWT 7.1 use the sample 1951-2010. PWT 9 use the sample 1951-2014. Data downloaded from http://www.rug.nl/ggdc/productivity/pwt/. (Accessed 15 March 2017) For PWT9, the GDP per capita growth series is calculated as the log growth rate of national accounts real GDP divided by population; using PWT9 variable names: $ln(rgdpna_t/pop_t) - ln\,(rgdpna_{t-1}\,/pop_{t-1})$. For PWT7.1: our GDP per capita variable is *rgdpl:* Real GDP per capita (Constant Prices: Laspeyres). We generate $growth_t=\ln(rgdpl_t)-\ln(rgdpl_{t-1})$

World Bank World Development Indicators Growth Data

We use GDP per capita growth (annual %) (NY:GDP.PCAP:KD.ZG). Data can be downloaded from: http://data.worldbank.org/data-catalog/world-development-indicators
(accessed 17 March 2017) We convert ppt actual growth rates G into log growth rates: growth=log(1+G/100)

Data Sample and Cleaning

WDI growth data was only available starting in 1961 (and so the sample runs 1961-2014, with an end date to match that in PWT9). We drop observations where |growth|>0.4 as described in the text, as well as Kuwait in 1990/91, Liberia and Myanmar before 1970 (see Section 3 and Appendix table 1B).[54] We drop countries with less than 30 years of growth data (combined across all our data sources).

[53] To merge 3-letter country isocodes and Correlates of War country codes we used Andreas Beger's crosswalk (http://myweb.fsu.edu/ab05h/research.html#dofiles). We thank Andreas Beger for making this publicly available. An earlier version of this paper used leader data from Jones and Olken (2005) – we thank Ben Jones and Ben Olken for sharing their data with us.

[54] We drop data for Myanmar before (and including) 1970 as the measures of GDP per capital trebles in 1970, which might be due to a change in methodology that make measures GDP before and after 1970 comparable.