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

William Easterly and Steven Pennings[1]

New York University and World Bank

This version: 25 May 2017
Preliminary – comments welcome.

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 (controlling for commodity prices, regional business cycles, and country effects). While we do sometimes find sizable growth contributions of celebrated "benevolent autocrats", we also find that (i) they are regularly outranked by other less celebrated leaders and (ii) the ranking and contributions of leaders is often not robust across growth datasets. Moreover, we find that even in world where leaders do affect growth, the average growth rate during a leader's tenure is mostly uninformative about that leader's actual growth contribution. Depending on the dataset and methodology, we find that that measured least squares leader contributions and unobserved leader effects can vary just as much in democracies as autocracies.

---

[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, Angus Deaton, and participants in the NYU Development Seminar, NYU Development Research Institute Annual Conference, NEUDC and Mid-west political science conference. Jorge Guzman provided assistance with data. A previous (and substantially different) version of this paper was circulated under the title "How Much Do Leaders Explain Growth? An Exercise in Growth Accounting" Web: http://williameasterly.org/ and https://sites.google.com/site/stevenpennings/

Section 1: Introduction

Debates in development on democracy versus autocracy are influenced by some case studies of famous autocrats such as Deng Xiaoping, Lee Kuan Yew, or Paul Kagame presiding over high growth episodes. Both popular and academic discussions give much credit to these "benevolent autocrats" for the growth outcomes. 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] "Singapore…has managed through benevolent dictatorship to produce a high quality of material life for its citizens, albeit without many of the freedoms that others hold dear." Bueno de Mesquita and Smith (2011). 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. "Highly centralized societies … may get a preceptor like Lee Kwan Yew of Singapore or a preceptor like Idi Amin of Uganda…" Sah (1991). 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 to do 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] Although autocracy is a gamble, the upside often gets the emphasis -- "visionary leaders … in autocratic [governments] need not heed legislative, judicial, or media constraints" (Becker 2010).

Despite the importance of the question of how or whether to attribute growth to leaders, there has been surprisingly little formal quantitative analysis, a gap this paper aims to fill. The most notable exception has been Jones and Olken (JO) (2005) who find that economic growth changes (in either direction) when a leader dies unexpectedly in office (such as by illness or accident) — with the results significant for autocracies but not for democracies. These results seem to confirm the previous stylized facts that autocrats have a higher variance of growth outcomes than democrats. JO admirably addressed causality by using the exogeneity of accidental deaths. The JO result is so important that it needs replication, especially since the sample of leader deaths was of necessity very small. For example, the non-result on democratic leaders may have been due to low power in a small sample of 22 leaders. Because particular leader cases are important in influencing both the democracy debate and aid decisions, we also

---

[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).

want a method that goes beyond JO to evaluate ANY particular leader for the magnitude of their positive or negative growth effect.[5]

In this paper, we ask: how much economic growth can we *quantitatively* attribute to a particular leader, based on the average growth rate during their tenure (as well as other observables)? In the literature on the importance of autocrats for growth, there is an implicit belief that this proportion is close to one, or at least well above zero. Using this metric, we want to know who are the best (and worst) leaders for growth, and what is the size of their contribution? This is a difficult problem due to all the other factors influencing growth, such as noise (Easterly et al 1993), country-specific factors, and other determinants.

A key insight is that 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 voluminous literature (eg Kane and Staiger 2008, Chetty et al 2014 among many others). To our knowledge, we are the first to make this connection.[6] Here the national leader represents the schoolteacher, and the average test scores of past classes taught by the teacher represents the average growth rate during the leader's tenure.[7] In the schoolteacher VA literature, authors first remove observable determinants of student's test scores (such as demographic factors) via a regression, and then adjust the past average residual test scores for the signal-to-noise ratio. As past test scores are noisy measures to true performance – as teachers only teach a limited number of students --- they need to be "shrunk" towards zero, with the size of the adjustment known as the reliability or shrinkage factor. Teacher value added estimates constructed in this way yield the best ("least squares") predictor of the teacher's true contribution to test scores. The reliability/shrinkage factor is constructed using estimates of variance components (variation in true teacher quality or noise) as this determines the signal-to-noise ratio.

In this paper, we adapt teacher VA approach to the many important differences in the leader-growth context (Section 2) to calculate best (least squares) estimates of the contribution of each leader to growth. The largest departure from the teacher VA literature is on data quality, heterogeneity and sample size. Economic growth is much harder to measure than a test score, and so we report results using both Penn World Tables 9 growth data and Penn World Tables 7.1 (we sometimes also use the World Bank's World Development Indicators) – with results often differing depending on the dataset used (see Section 3 for a description of data). Estimating the variance components to create the reliability/shrinkage factor is difficult due to heterogeneity in country-level noise and leader tenure, and so in our main results we use a method that adjusts for unbalanced panels and test its performance using Monte Carlo simulations (Section 4). All of this is made more challenging by the fact that most leaders are only in power for a few years, whereas teachers teach hundreds of students. A consequence is that we have 7000 observations, whereas Chetty et al (2014) have 7 million.

Two other differences are on observables and country effects. With leaders, there are few *exogenous and observable* determinants of growth that can be controlled for. We remove the effect of commodity prices and regional business cycles (via continent X year dummies) to produce growth residuals, but these have much less

---

[5] Other contributions include: Besley et al (2011) who extend JO's findings to show a positive effect on growth of the leader's educational attainment. These authors also find evidence of leader deaths on growth (across their whole sample), but don't report an estimate of the contribution of leaders to growth. Meyersson (2016) examines coups, and finds that while successful coups in autocracies have an imprecise effect on growth, coups in democracies tend to reduce growth rates. In a recent study, Blinder and Watson (2016) discuss average growth under US Presidents, and find that the average growth rate is higher under Democrats than under Republicans.

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

[7] A caveat of this approach is that it misses potentially important *lagged* effects of teachers/leaders on test scores/growth.

effect than demographic factors in the teacher VA literature. Related, we adjust for the impact of country effects on growth (for example, geography, history or culture), using the average growth rate under *other* leaders in the same country, which is shrunk based on its own signal to noise ratio. There is no direct analogy of country effects for teacher VA. Because leaders only lead one country, country effects can't be controlled for by regression.[8]

Our first result is that even using a model where leaders do affect economic growth, the average growth rate under a certain leader is mostly uninformative about that leader's true contribution to growth. On average this means that only one quarter to one eighth of an increase in the average growth rate translates into an increase the least-squares leader contribution (Section 6). This suggest that even one takes Jones and Olken's (2005) estimates of the distribution of true leader quality as given (ours are often not that different, depending on the dataset used) even a relatively long average says surprisingly little about how good or bad a particular leader is for growth.

Among the set of possible best and worst leaders, the average growth rate during the leaders' tenure is almost *completely uninformative* about the least squares ranking of the leaders (Section 7). This is because countries with extremely good or bad leader growth averages tend to have more noisy growth, which results in more shrinkage towards zero (this is also true because many of these leaders have short tenures). The leader growth average also misses country effects – producing strong growth in a country which always grows strongly is less of an achievement than producing growth in one that is beset by bad fundamentals (controlling for regional growth can also be important). This means that many of the famous 'benevolent autocrats" which supposedly produced high growth rates (albeit at the cost of political repression) turn out not to be as good for growth as their usual narratives suggest. While there do seem to be a few good-for-growth autocrats that are robustly identified – such as Khama (Botswana), Chun Doo Wan (South Korea) and Medici (Brazil) – it is difficult a priori to work out the best from the rest (and even the best contribute much less than the average growth rate during their tenure). Moreover, the best and worst leaders (even using our least squares methodology) also vary across datasets, which make it difficult to make strong statements regarding leader quality.

Our second result is that we fail to robustly confirm the common finding that the contribution of autocratic leaders to growth is larger than that for democratic leaders in both directions (i.e autocratic leaders are a risky bet). This finding is usually justified in the literature that autocratic leaders facing fewer constraints on their power. Like others in the literature, we find that the average growth rate is more variable under autocrats than democrats (a SD 1.5 times as large). However, much of this difference is just short-run noise and country-level factors. When we calculate the standard deviation of the least-squares leader effects, the standard deviation (across leaders) is *smaller* under autocracies than under democracies using PWT 7.1 data, though the reverse is true for PWT9 data (Section 6). Related, we also find that the unobserved standard deviation of leader quality is the same for autocrats as democrats using the PWT7.1 data -- around 1% for each (Section 5). Using PWT9 data, we find that the unobserved leader SD is larger for autocrats than democrats (1.5% vs 1%), however, this is not robustness across methodologies.[9] Using a standard random effects estimator of the true leader SD --- which sometimes performs better in Monte Carlo tests –the autocratic contribution is actually close to zero. The sensitivity of results to the dataset used is in line with Johnson et al (2013), who find that when Jones and Olken's (2005) results are re-estimated using PWT 6.2 data, it is democratic leaders rather than autocratic ones who affect growth (the opposite of the original result which was calculating using PWT 6.1). Combined, these results suggest the view of growth driven by "unconstrained" good or bad autocrats is overly simplistic, perhaps because autocratic leaders find it difficult to know how to boost growth, and to implement pro-growth policies in face other political constraints.

---

[8] This also precludes many identification strategies used in the teacher VA literature. An exception is Yao and Zhang (2015), who use the fact that Chinese city mayors switch cities to estimates the effect of mayors on growth (they find mixed results).

[9] Overall the leader SD is around 1.5%, which is almost identical to what Jones and Olken (2005) find.

Section 2: Model and Methodology

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

In the academic literature and in policy discussions, leaders are often attributed the average growth during their tenure, as discussed above. Even if we give leaders as much credit for growth as possible, there are still three problems with this approach. First, the random *idiosyncratic* 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. This means a good string of good (or bad) growth rates under a leader are attributed to the good (or bad) policies of a leader, when often they are just good (or bad) luck. 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.[10] Finally, there are *other supranational forces* that can affect growth, such as commodity prices or global/regional business cycles.

This simple 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), time varying global or regional business cycle $\bar{g}_t$, a vector of observables like commodity prices in X, a country-specific component (μ_c) and idiosyncratic error (ε_{ict}) component for a panel of leaders:

$$(1)\quad g^*_{ict} = \bar{g}_t + X\beta + \mu_i + \mu_c + \varepsilon_{ict}$$

$\bar{g}_t + X\beta$ are observable, which means that we can control for them, leaving the growth residual g_{ict} (Equation 2). g_{ict} then depends on three unobserved random variables μ_i , μ_c , ε_{ict}, which from which the country draws $\mu_c \sim (0,\sigma_c^2)$, each leader draws $\mu_i \sim (0,\sigma_\mu^2)$ and for each period $\varepsilon_{ict} \sim (0,\sigma_{c\varepsilon}^2)$, with μ_i , μ_c and ε_{ict} being independent (and also serially uncorrelated). We assume ε_{ict} is independent across years and countries, but we allow it to 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, which suggests these data are particularly unreliable. Each leader is in power for T_i years and N_c is the total number sample length for country c.

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

Note that we are intentionally modeling growth to be as favorable as possible to the practice of attribution of growth to leaders. We give leaders full credit for *all* growth during their tenure except for that due to observable international factors (international business cycles and commodity prices), country effects and iid shocks. For example, we rule out anybody else in government other than the leader having any effect on growth (bad luck for finance ministers and central bank governors). 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.[11]

The average growth rate under leader is usually attributed in the literature to the strength (or weakness) of that leader. We define *the leader residual growth average*, $\bar{g}_{ic}$, as the average of the *residual* growth rate of leader *i* in

---

[10] 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).

[11] 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).

country *c* with tenure T_i , after we remove observables such as commodity prices or the international 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 (which we denote –*i*) after removing for observables. This is going to be helpful to distinguish between country effect and individual leader 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 as regressions on observables: (i) time by region fixed effects to capture the international business cycle and (ii) on a commodity price index for the country, and then producing an estimate of the residuals g_{ict}. In practice it usually turns out the international business cycle and commodity price fluctuations are of second order importance, and the largest impact is to remove the mean worldwide growth rate (which is around 2%) from the leader growth average

Note that we can't remove the country effect using a regression of growth on country dummies, because for specific countries we cannot distinguish the country effect from a string of leader effects (which are also unobservable).[12] This problem is particularly striking for a number of countries have had only a few leaders, or had one leader who was in power for most of the sample. For example, growth was strong under Goh Chok Tong, but below that under Lee Kuan Yew, who led Singapore for 30 years. If Mr Lee was a great leader (a high μ_i), then is also possible that Mr Goh was a good leader, and that Singapore has a modest μ_c overall. But if we impose that μ_c is high for Singapore via a regression on country dummies– for example by setting it equal to Singapore's high average growth rate – then we are also imposing that Mr Lee was at most a okay leader, and Mr Goh was a *bad* leader (a low μ_i).[13]

<u>Definition of problem</u>

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 also the average growth rate residual under *other* leaders in the same country $\bar{g}_{-ic}$. "Best" here is the minimum least squares error, as commonly used for evaluating forecasts (in the empirical work and Monte Carlo simulations we report the Root Mean Squared Error RMSE). Ideally we would like an unbiased estimator, though we don't impose it. We also restrict the model such that the best 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 as a proxy for the country effect (which is then subtracted --this proxy and our confidence in it varies

---

[12] Note that we do estimate the variance of country effects (across countries) using this methodology, which produces slightly upward biased estimates of the true country effect. But it is much more damaging for estimating the effect of individual leader contributions

[13] This also means that if a country is prone to producing good leaders, we would identify that as a country effect and we would not give credit for that tendency to individual leaders.

depending on the leader – see discussion below). ψ is the *reliability factor* (sometimes called the *shrinkage factor*) which adjusts downward the adjusted leader growth to minimize the error variance.[14]

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

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

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

The optimal $\hat{\gamma}$ is just the population regression estimate of $\bar{g}_{ic}$ on $\bar{g}_{-ic}$ (Equation 7), where σ_c^2 is the population variance of the country effect, σ_e^2 is the variance of the iid error term (which could vary by country), and σ_μ^2 is the variance of the leader effect.[15] 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}).

(7) $\hat{\gamma} = \dfrac{E(\bar{g}_{ic}\bar{g}_{-ic})}{E(\bar{g}_{-ic}{}^2)} = \dfrac{\sigma_c^2}{\sigma_c^2 + \dfrac{\sigma_e^2}{N-T_i} + \dfrac{\sigma_\mu^2 T_{-i}}{N-T_i}}$

If σ_c^2=0 then $\hat{\gamma}=0$, then 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. In contrast, when $\sigma_e^2 = \sigma_\mu^2$=0 and σ_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) the country sample size (N) is long relative to the tenure of an individual leader (T_i), 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.

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

The estimate of the *reliability factor* (also known as 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 average residual is a perfect signal of the leader's effect on growth.

However, if $\sigma_e^2 > 0$ then the optimal reliability factor $\hat{\psi}$<1 because the leader growth average includes noise due to the idiosyncratic shocks to growth ε_{it}. $\hat{\psi}$ 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

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

[15] This is not by construction, but because $E(\mu_i\bar{g}_{-ic}) = 0$.

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 reliability 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). In the rest of the paper, we estimate the variance components of Equations (6) and (7) and make our best estimate of the contribution of each leader to growth.

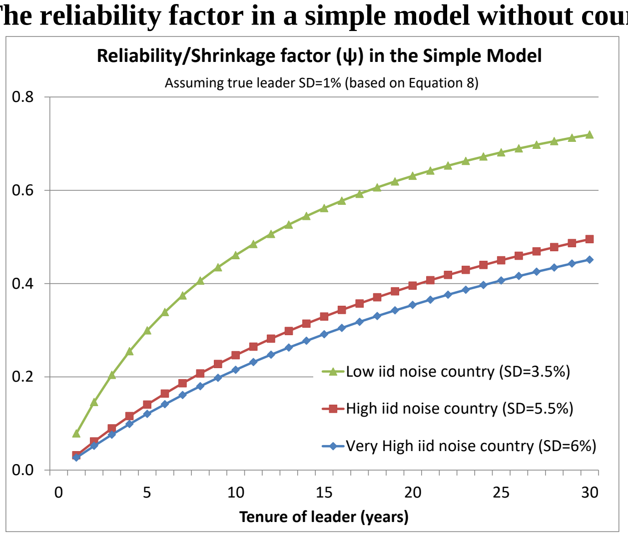
The reliability factor in a simple model without country effects

In our full model (Section 6) the most important drivers of the reliability/shrinkage factor are the variance of iid noise and the length of leader tenure. To illustrate this, we assume that the *unobservable* standard deviation of the leader effect is 1% across all leaders (similar to results in Section 5 using PWT 7.1 data), and there are no country effects ($\sigma_c^2 = 0$), which means that Equation 8 can be applied directly to leader growth average residual, and simplifies to:

$$(9)\quad \hat{\psi}_{Simple} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_e^2 \;/T_i}$$

In Figure 1 we plot the reliability factor vs leader tenure for low iid error variance (σ_e =3.5%) and high iid error variance (σ_e =5.5%), and very high iid error variance (σ_e =6%). We will see that these numbers roughly correspond to the variance under democracy, all leaders and autocracy, respectively. One can see that, in general, the average growth rate during a leader's tenure is relatively uninformative about the size of the true (but unobservable) leader effect μ_i (a low reliability factor ψ). However, it is much less informative for very high iid error countries than lower iid error countries. For example, for a leader of 5 years tenure (close to the average), the reliability factor is *two and a half times larger* in the low iid noise countries than in very high iid noise countries, with a reliability estimate of around 0.12 in the latter. We will see that autocracies tend to have much more noisy growth processes (very high SD of around 6%) than democracies (low SD around 3.5%), which is why the average growth rate under autocrats is often a particularly *unreliable* estimate of the true leader contribution. This is particularly true for autocrats with short tenures. Country effects (included in the full model below), reduce reliability estimates for leaders with very long tenures because it becomes hard to distinguish between the contribution of the leader and the contribution of the country.

Figure 1: The reliability factor in a simple model without country effects



Relation to Teacher Value Added literature

The teacher value added literature seeks to answer a similar question to the one we ask here: replacing teachers with leaders and test scores with economic growth (though there are some important differences).[16] The methods in the teacher VA literature have generally been quite successful in producing estimates of teacher VA.[17] For example, Kane and Staiger (2008) conduct an experiment with the random assignment of teachers (which unfortunately we can't do for leaders), and found that non-experimental teacher VA estimates were unbiased, conditional on controlling for past test scores, and quite accurate after controlling for mean classroom characteristics. Chetty et al (2014) find that teacher VA estimates are unbiased forecasts of achievement with respect to parental characteristics from tax records, not used in the construction of the VA estimates. They also find that VA estimates predict changes in test scores in event studies where teachers change schools.

The approach here follows the methodology in the teacher value added literature such as Kane and Staiger (2008), and more recently Chetty et al (2014) (among many others).[18] The teacher VA literature first removes exogenous observable determinants of individual student test scores – the most important being prior year test scores for the student, but also demographics characteristics of the student and average demographics characteristics of the school/classroom. In Kane and Staiger (2008), these factors also reduce the SD of true teacher estimates. In the case of national leaders, economic growth is much more difficult to explain than students' performance, and as a result observable covariates are much less important. Crucially, there are also few exogenous determinants of economic growth, and also there is a much smaller sample to estimate relationships (around 7000 growth-year observations, whereas Chetty et al have 7 million observations).

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 reliability/shrinkage factor in Equation 9 is the same as Equation 9 in Chetty et al (2014) where the student error corresponds to annual GDP growth error.[19] The country effect doesn't have a direct analog in an education context, but would be roughly equivalent to the case where there was a time-invariant class effect under multiple teachers. For example, in a small school the same group of students might have been in a class together for several years and so class-specific factors (how the class members get along) can affect test scores under multiple teachers. The teacher VA literature doesn't explicitly control for this, perhaps because it is of second-order importance in their context.[20]

Finally, papers in the teacher VA literature estimate the true variance of teacher VA from the *covariance* of test score residuals over time (under the same teacher). In our case, this would deliver a combination of the leader and country effects, though the latter could in principle be removed using the covariance of growth across leader transitions (which is an interesting area for future research). As described below, we instead estimate the true leader

---

[16] Kane and Staiger (2008) summarize the teacher VA literature as answering a very specific question: suppose the students in a particular classroom had teacher A rather than teacher B, how much different would their average test scores be? On average, test scores should increase by the difference in the two teacher's value added. Here we want to create the leader value added such that if we replaced leader A with leader B, per capita economic growth would increase by the leader estimate.

[17] 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.

[18] Chetty et al (2014) broadly follow Kane and Staiger's (2008) methodology for estimating teacher effects, with the exception that they allow for drift in teacher quality (more recent average scores are a better guide for VA than less recent ones). As we can only estimate the leader contribution at low frequency, drift in leader quality is not an issue for leaders.

[19] Equation 9 in Chetty et al is the same as Equation 5 in Kane and Staiger (2008) with a constant classroom size. Alternatively, one could assume an infinite number of students per classroom, and then the iid growth error is like the classroom effect.

[20] Chetty et al (2014) control for cubics in prior year class means, which might partially control for some of this.

SD using the difference between the average growth rate during the leader's tenure and the mean iid error (after removing the country effects).[21] The teacher VA literature generally doesn't test their methods using Monte Carlo simulations (with a few notable exceptions, such as Kinsler 2012).

Section 2.2: 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 $\sigma_c^2, \sigma_\mu^2, \sigma_\varepsilon^2$. Moreover, σ_μ^2 is of general 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 – although it is not as straightforward as that.

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 using Equation (10) where $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_{\mu1}^2$ can be estimated using standard variance formulas in Equation (11) and (12) (formulas provided for balanced panels).

(10) $$\hat{\sigma}_\mu^2 = \hat{\sigma}_{\mu1}^2 - \hat{\sigma}_\varepsilon^2 / T$$

(11) $$\hat{\sigma}_{\mu1}^2 = \left[\frac{1}{N}\sum_{i=1}^{N}(\bar{g}_i - \bar{g})^2\right]$$

(12) $$\hat{\sigma}_\varepsilon^2 = \frac{1}{N_L(T-1)}\sum_{i-1}^{N_L}\sum_{t=1}^{T}(g_{it} - \bar{g}_i)^2$$

It is possible for $\hat{\sigma}_{\mu1}^2$ to be negative if $\hat{\sigma}_{\mu1}^2$ is small and so the estimator replaces negative estimates with zero (i.e. $\hat{\sigma}_\mu^2 = \max(0, \hat{\sigma}_{\mu1}^2 - \hat{\sigma}_\varepsilon^2 / T)$), with the Monte Carlo studies finding this not being a serious problem (Baltagi 2005 p18).

We use two variations of Equation (10) to generate feasible estimates of the true leader effects $\hat{\sigma}_\mu^2$: the standard random effects estimator (which we label RE after the Stata command used to estimate it), and the second is similar but includes a small sample correction for unbalanced panels from Baltagi and Chang (1994) (which we label SA, again, after Stata's command).[22] The methods are identical for balanced panels. Our panel of leaders is very unbalanced, so one would think that the SA method would be preferred, but that is not always the case.

---

[21] Kane and Staiger (2008) estimate the individual student error component as the variance of the test score residual less the class average, and class effect as any remaining variation (Equations 2-4 in Kane and Staiger 2008).

[22] Both of these methods use Swamy-Arora's approach to calculate residuals, which involves calculating $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_{\mu1}^2$ using the residuals from two regressions: $\hat{\sigma}_\varepsilon^2$ is calculated from the residuals of a within regression (only time variation) and $\hat{\sigma}_{\mu1}^2$ is calculated using a between regression (only cross-sectional variation). Baltagi and Chang (1994) show their unbalanced panel small sample adjustment show performs well in Monte Carlo simulations. The methods are implemented in Stata using xtreg, re (default) and xtreg, sa (with unbalanced panel correction). See the Stata manual, Baltagi (2005) and Baltagi and Chang (1994 for further details).

One way that we *cannot* calculate the variation of leader effects is just by calculating the standard deviation of the leader growth average (or leader fixed effects), as in Equation 11 – and as reported when the xtreg, fe command is used in Stata. One of the main points of this paper is that the leader growth average is only weakly informative about the true leader effect because the presence of iid noise only averages out slowly, which means that the variance of the average growth rates will be substantially biased upwards as a measure of leader quality variation, as in Equation (13) (shown for the simple model).

$$E\hat{\sigma}_1^2 = \sigma_\mu^2 + \frac{\sigma_\varepsilon^2}{T} > \sigma_\mu^2 \qquad (13)$$

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.[23] There are no econometric issues in estimating the iid error (Monte Carlo evidence suggests it is in fact estimated very accurately). The country effect is estimated by adding country dummy variables to the random effects regression of 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 where $\hat{\sigma}_{c1} = \sqrt{\sigma_c^2 + \sigma_\mu^2/(N_c/T) + \sigma_e^2/N_c}$. 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 in Section 4 via Monte Carlo simulation). In principle, one could correct for this using a similar approach as in Equation (10) $\hat{\sigma}_c = \sqrt{\sigma_{c1}^2 - (\sigma_\mu^2/(N_c/T) + \sigma_e^2/N_c})$, which is an interesting extension for future work. Monte Carlo estimates suggest this adjustment could increase the accuracy of the country effect estimates

Section 3: Data description and data quality (or lack thereof)

Section 3.1: Data Sources

In order to estimate the size of leader effects we need data on leaders, growth and a measure of whether each country is a Democracy or Autocracy. Data which leaders are in power is taken from Archigos 4.1 dataset (Goemans et al 2009).[24] Following Jones and Olken (2005), we use the log growth rate: $\ln(Y_t)$- $\ln(Y_{t-1})$, where Y_t is real per capita GDP.

We use data on real GDP per capita growth from several sources: in the body of the text we the Penn World Tables (PWT) version 9 (the latest version at the time of writing, Feenstra et al 2015) and PWT 7.1 (the latest at time of the first draft of this work; Heston et al 2012). We also use real per capita GDP growth from the World Bank's World Development Indicators (WDI) as a cross-check, with results mostly presented in the appendix.[25] Also see the Appendix for further details on data sources and construction.

---

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

[24] 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. The Archigos 4.1 dataset includes leaders up until 2015, which substantially increases the size of our dataset. The Archigos dataset also covers more countries than Jones and Olken (2005). Over a common sample of years and countries, the Jones and Olken (2005) and Archigos datasets are almost identical (unsurprisingly).

[25] 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).

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 so might be alternatively named "established democracies". However, it is only 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 -4). 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. For the rest of the paper we refer to non-democracies as autocracies.

Section 3.2: Descriptive Statistics

We have 136 countries for which we have growth, leader and polity data, of these about 20% are democracies (see Appendix Table 1 for the 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.[26]

Table 1 shows the basic descriptive statistics. We have around 7000 observations (6000 for WDI) and 1100 leaders (1000 for WDI). Average per capita growth is about 1.9% per annum, and is higher on average in democracies than autocracies. The unconditional variance of growth is much higher for autocratic countries than democratic ones.

Table 1: Growth Descriptive Statistics

	Mean	SD	Obs	Leaders	Tenure
A. All					
PWT 9	1.91%	6.28%	7214	1168	6.2
PWT 7.1	1.89%	6.81%	6764	1115	6.1
WDI	1.86%	5.87%	6275	1020	6.2
B. Autocracies	Mean	SD	Obs	Leaders	Tenure
PWT 9	1.77%	6.91%	5598	824	6.8
PWT 7.1	1.75%	7.45%	5272	796	6.6
WDI	1.76%	6.43%	4966	748	6.6
C. Democracies	Mean	SD	Obs	Leaders	Tenure
PWT 9	2.40%	3.20%	1616	344	4.7
PWT 7.1	2.42%	3.72%	1492	319	4.7
WDI	2.24%	2.90%	1309	272	4.8

Notes: Descriptive statistics calculated using all observations (including outliers)

Section 3.3 Outliers

Per capita 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. Other things equal, a growth observation 5 percentage points above the mean has 100 times the weight of one 0.5 percentage points above the mean. Things get worse for very extreme observations: a growth rate 50 percentage points above the mean has 10000 times the weight of one 0.5% above the mean. These extreme observations do exist, for example, for countries entering or exiting civil wars. By this logic, a couple of coincidental leader

[26] 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.

transitions around times of civil wars or other extraordinary events can completely change our results, and overturn the evidence of thousands of other observations.

We take a very conservative definition of outliers – log growth of more than 40% (in absolute value) in particular year – and drop these from our main results. There are only around 12-18 outliers per dataset for the 6000-7000 observations. The individual observations dropped are listed in Appendix Table 1a. We also drop Kuwait in 1990 and 1991 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).

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 of 1991, the Rwandan genocide of 1994 (and rebound in 1995), the Lebanese civil war in the late 1970s and early 1980s and the first Liberian civil war around the early 1990s (and rebound in 1997 with peace). 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 25% (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.

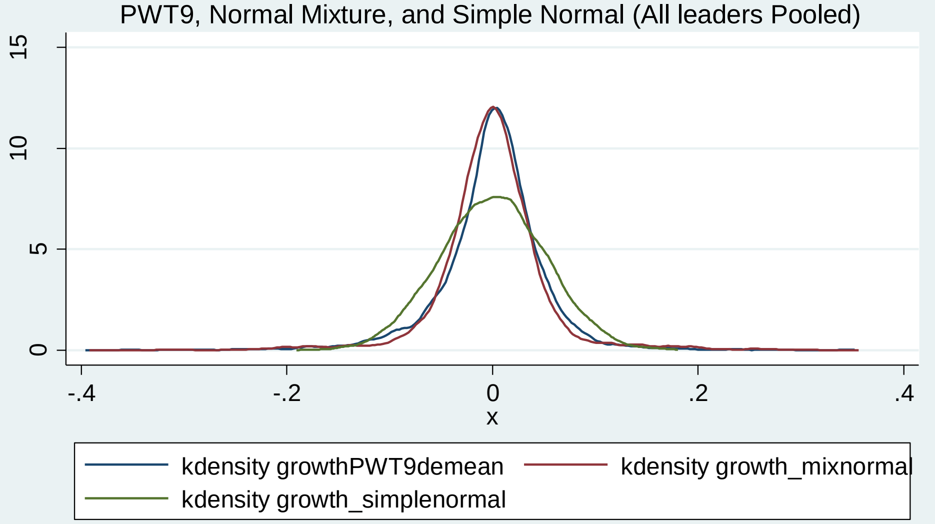
Section 4: Monte-Carlo Results

Section 4.1: Monte Carlo evidence on estimators of variance components

To evaluate the performance of each methodology we perform a Monte Carlo simulation of annual growth rates as in Equation 1 and 2 (Table 2), with either real or nonexistent country effects. In each iteration we draw a leader effect (μ_i), a country effect (μ_c, equal zero if there are no country effect), and an iid error (ε_{ct}) to generate growth data, combined with the actual leadership structure from Archigos dataset (that is we don't model commodity prices or regional business cycles). In panel A, we estimate a simple model without any country dummies, and in Panel B we estimate the full model with country dummies.

Figure 2: Kernal Density of Growth (SD=5.25%)

PWT9, Normal Mixture, and Simple Normal (All leaders Pooled)



Notes: Figure reports kernal density estimate of PWT9 real per capita growth rates (demeaned, excluding 40% outliers and LBR), a univariate norman distribution with the same standard deviation (5.25%), and the a mixture of 2 normals chosen to match the Std Dev and also the Kurtosis of 10.

A challenge here is that growth data is not normally distributed -- even when extreme outliers have been removed (tests of normality are rejected at the 1% level). The primary problem is that there is excess Kurtosis – too many extreme positive and negative growth observations relative to a normal distribution. In a normal distribution, the Kurtosis is 3, in raw PWT9 growth data it is around 50, and after removing 40% outliers (and Liberia) as above, Kurtosis is around 10. One can also see this in the kernel density plot in Figure 2, where the blue line is the distribution of growth in the data (demeaned and without outliers), which is more "peaked" around zero relative to a normal distribution with the *same* standard deviation (green line), and with more extreme observations above 20% in absolute value. To match the distribution of the actual data, we form a distribution as a mixture of normals. With probability of around 85%, we sample from a normal with a SD of 3% (to capture the "peak" of the distribution with growth less than 7% in absolute value) and with 15% probability we sample from a normal distribution with an SD of 11%, calibrated to generate kurtosis of 10 as in the data. The resulting distribution of mixture of normals is shown in red in Figure 2, and closely matches the distribution of growth data. We draw growth rates from a distribution analogous to this– but constructed for autocratic and democratic sample separately.

Table 2: Monte Carlo Estimates of Variance Components (True: sd(leader)=1.5%)

	Panel A: All Leaders (Pooled sample of democracies and autocracies below)									
	Simple Model (no country dummies)				**Full Model (with country dummies)**					
	SA-Method		**RE-Method**		**SA-Method**			**RE-Method**		
	SD(leader)	**sd(iid)**	**SD(leader)**	**sd(iid)**	**SD(leader)**	**sd(iid)**	**sd(CE)**	**SD(leader)**	**sd(iid)**	**sd(CE)**
No Country Effects	1.40%	5.29%	0.99%	5.29%	1.69%	5.29%	1.03%	1.06%	5.29%	1.04%
True sd(CE)=0	[0.13%]	[0.1%]	[0.56%]	[0.1%]	[0.16%]	[0.1%]	[0.07%]	[0.56%]	[0.1%]	[0.07%]
Country Effect Pvalue:*							67.52%			24.97%
With Country Effects	2.04%	5.29%	1.81%	5.29%	1.69%	5.29%	1.81%	1.05%	5.29%	1.82%
True sd(CE)=1.5%	[0.12%]	[0.09%]	[0.35%]	[0.09%]	[0.15%]	[0.09%]	[0.11%]	[0.56%]	[0.09%]	[0.11%]
Country Effect Pvalue:*							0.00%			0.00%
	Panel B: Autocrats (iid error SD 5.7%, Kurtosis=9)									
	Simple Model (no country dummies)				**Full Model (with country dummies)**					
	SA-Method		**RE-Method**		**SA-Method**			**RE-Method**		
	SD(leader)	**sd(iid)**	**SD(leader)**	**sd(iid)**	**SD(leader)**	**sd(iid)**	**sd(CE)**	**SD(leader)**	**sd(iid)**	**sd(CE)**
No Country Effects	1.50%	5.69%	1.41%	5.69%	1.89%	5.69%	1.10%	1.52%	5.69%	1.10%
True sd(CE)=0	[0.15%]	[0.11%]	[0.61%]	[0.11%]	[0.19%]	[0.11%]	[0.08%]	[0.6%]	[0.11%]	[0.08%]
Country Effect Pvalue:*							79.44%			52.10%
With Country Effects	2.13%	5.70%	2.07%	5.70%	1.89%	5.70%	1.86%	1.50%	5.70%	1.87%
True sd(CE)=1.5%	[0.15%]	[0.11%]	[0.42%]	[0.11%]	[0.19%]	[0.11%]	[0.13%]	[0.6%]	[0.11%]	[0.13%]
Country Effect Pvalue:*							0.00%			0.00%
	Panel C: Democrats (iid error SD 3.2%, Kurtosis=7.3)									
	Simple Model (no country dummies)				**Full Model (with country dummies)**					
	SA-Method		**RE-Method**		**SA-Method**			**RE-Method**		
	SD(leader)	**sd(iid)**	**SD(leader)**	**sd(iid)**	**SD(leader)**	**sd(iid)**	**sd(CE)**	**SD(leader)**	**sd(iid)**	**sd(CE)**
No Country Effects	1.49%	3.20%	1.46%	3.20%	1.57%	3.20%	0.64%	1.49%	3.20%	0.64%
True sd(CE)=0	[0.14%]	[0.11%]	[0.33%]	[0.11%]	[0.16%]	[0.11%]	[0.1%]	[0.33%]	[0.11%]	[0.1%]
Country Effect Pvalue:*							56.02%			51.22%
With Country Effects	2.09%	3.20%	2.06%	3.20%	1.57%	3.20%	1.61%	1.48%	3.20%	1.61%
True sd(CE)=1.5%	[0.2%]	[0.11%]	[0.29%]	[0.11%]	[0.15%]	[0.11%]	[0.23%]	[0.29%]	[0.11%]	[0.23%]
Country Effect Pvalue:*							0.0%			0.0%

* P-value of test country effects (CE)=0. Note: Table presents monte carlo estimates of leader effects, where the real country X leader structure is used, but leader effects are drawn from a normal distribution with true SD 1.5%. A successful method uncovers the "true" parameter of the leader effect of 1.5%. In the left panel the method has no country dummies, whereas on the right panel is the method with country dummies to detect country effects. iid errors are drawn from a mixture of normals, to replicate excess kurtosis in actual growth data (different values for autocracies and democracies). The All Leaders sample is pooled version of the democrat and autocrat samples. See Appendix Table 2 for an all-leaders SD where all leaders have the same SD iid error. Standard deviations of bootstrap sample (across replications) are reported in brackets.

Neither the SA method (with an unbalanced panel adjustment) nor standard random effects (RE) estimator is unambiguously better at estimating the true leader SD of 1.5% in Table 2. Focusing on the most relevant case of the full model and a world with true country effects (bottom RHS of each panel), the RE effects estimator does extremely well in separate autocrat/democrat samples (Panel B and C), but is substantially downward biased (by 0.45ppts) in panel A with heterogeneous error variances (a mix of democrats and autocratic countries). In Appendix Table 2, we show that in the pooled sample when the error variance is homogenous, the RE method uncovers the true leader SD almost perfectly.

SA has the opposite problem: it is consistently upward biased; with the bias being relatively small at around 0.2-0.3% in both pooled and autocratic leader samples (the bias is much smaller for democrats). As this paper questions a large leader effect that has been the focus of the literature, we are more concerned with downward bias than upward bias and so prefer SA over RE (though we report the RE results in the appendix).

Country effects and the iid error

The full model (with country dummies) is about as accurate whether there are country effects or not (for both RE and SA methods). However, the simple model (without country dummies) only performs well in a world without country effects -- otherwise estimates of the leader effect are substantially biased upwards as the methods confuse a high country effect for a string of good leaders. Our estimates in the following section suggest that country effects are non-zero, and so we prefer the full model over the simple model.[27]

Conditional on there being country effects – as we find in the data – the estimates of country effects are slightly upward biased by both RE and SA methods. That is the country effect is estimated at 1.85%, which is upward biased by around 0.35ppts. The bias is mostly due to attributing the average of iid effects and leader effects to countries – the same issue as the upward bias in the SD of the leader growth average. However, because the number of observations per country is much larger than the number of observations per leader, the country effect bias is much less of a problem than the leader effect bias.[28]

All methods estimate the iid error variance very accurately.

Section 4.2: Monte Carlo evidence on estimators Least-Squares Leader Effects

Conditional on Equations (1) and (2) being the true model of growth, there are two possible sources of error in our least squares leader estimate: (i) our method for estimating the leader effect is either inherently biased or inaccurate (a high RMSE) and/or (ii) our estimates of the variance components $\{\sigma_u^2, \sigma_c^2, \sigma_e^2\}$ are inaccurate (which also depend on which method is used to calculate them). In this subsection we test these two effects using a sequence of Monte Carlo tests.

---

[27] Tests for country effects performed well: with a p-value of 0 when there were country effects and a p-value of >0.25 when there were no country effects, and so we confident of this assessment.

[28] For example, this formula suggests $\hat{\sigma}_c \approx \sqrt{\sigma_c^2 + \sigma_\mu^2/(N_c/T) + \sigma_e^2/N_c} \approx \sqrt{0.015^2 + 0.015^2/6 + 0.053^2/55}$=1.8%

Table 3: Monte Carlo Estimates of Leader Effects (500 reps -- All Leaders)

	PSI Mean*	PSI (SD)*	GAMMA Mean*	GAMMA (SD)*	Unbiased^	RMSE
	Panel A: Using Actual Variance Components					
No CE	0.31	0.19	0.00	0.00	1.00	1.25%
					[4.44%]	[0.03%]
With CE	0.28	0.16	0.74	0.11	0.99	1.28%
					[4.64%]	[0.03%]
	Panel B: Using SA Variance Components					
No CE	0.29	0.16	0.54	0.10	0.94	1.29%
	[3.83%]	[1.18%]	[3.52%]	[0.4%]	[15.45%]	[0.03%]
With CE	0.28	0.15	0.78	0.08	0.90	1.31%
	[3.58%]	[1.07%]	[2.38%]	[0.64%]	[13.17%]	[0.03%]
	Panel C: Using RE Variance Components					
No CE	0.17	0.10	0.60	0.08	2.09	1.36%
	[11.32%]	[5.86%]	[5.52%]	[1.15%]	[386.23%]	[0.08%]
With CE	0.16	0.09	0.81	0.06	10.80	1.36%
	[10.79%]	[5.35%]	[3.44%]	[1.35%]	[17585.78%]	[0.07%]

* Mean and SD across leaders (does not change with draws of growth). ^ Unbiased =1 Notes: Calculated using the actual leader dataset and SD(leader)=sd(CE)=1.5%, sd(iid)=5%. Note: Table presents monte carlo estimates of leader effects, where the real country X leader structure is used, but leader effects are drawn from a normal distribution with true SD 1.5%. A successful method uncovers the "true" parameter of the leader effect of 1.5%. In the left panel the method has no country dummies, whereas on the right panel is the method with country dummies to detect country effects. iid errors are drawn from a mixture of normals, to replicate excess kurtosis in actual growth data (different values for autocracies and democracies). The All Leaders sample is pooled version of the democrat and autocrat samples. See Appendix Table 2 for an all-leaders SD where all leaders have the same SD iid error. Standard deviations of bootstrap sample (across replications) are reported in brackets.

Monte Carlo simulations suggest that our least-squares estimates of leader effects are unbiased and have a root mean squared error of around 1.3% (Table 3), though the SA method (unbalanced panel adjusted) are more accurate than estimating leader effects using the RE method. The top panel of Table 3 reports Monte Carlo estimates assuming that we know the size of the population variance components $\{\sigma_u^2, \sigma_c^2, \sigma_e^2\}$ – and so ignores inaccuracies due to mismeasurement of variance components. The first column reports the mean reliability (ψ) calculated using Equation 8, which is around 0.3 based on the true values of the variance components. This means that on average, a leader growth (residual) average of 1% will become one of 0.3% after removing noise. The SD of ψ across leaders is quite large at 0.2, which reflects the differences in tenures (mostly), but also in the number of leaders per country, and whether the error variance is high (autocracies) or low (democracies). The third column shows the mean estimates of γ, the degree to which we adjust leader growth averages for the performance of other leaders in the same country. γ averages around 0.7 across leaders in the model with country effects, though is zero in the model without country effects. This estimate varies across leaders (with a standard deviation of around 0.1) due to variation in the cumulative tenure of other leaders in the same country (and the number of other leaders in the same country).[29]

The fifth column of Table 3 reports tests of unbiasedness, where $\lambda = 1$ implies unbiased least squares leader estimates. Estimates of λ from a regression of the true leader effect μ_i on the least-squares leader estimate $\hat{\mu}_i$ using simulated data as in Equation 14. One can see that in both models (with and without country effects), the Monte Carlo estimates of λ are close to one. Estimates of the standard error of $\hat{\lambda}$ (the standard deviation across MC draws) suggest that one could not reject a test of $\hat{\lambda} = 1$.

$$(14) \qquad \mu_i = \lambda \hat{\mu}_i + e_i \quad \text{where} \quad \hat{\mu}_i = \hat{\psi}(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})$$

Even if the leader estimates are unbiased, they could still not be very "accurate". We measure accuracy as the Root Mean Squared Error for each estimated leader effect (Equation 15, where L is the number of leaders). These

[29] The cumulative tenure of other leaders helps to average out iid noise, and a greater number of other leaders averages out their leader effects.

estimates are around 1.25% (with the RMSE error estimated quite accurately) --- even if we know the true variance components. It is worth noting that even though this is the minimum error as described in Section 1, there is still a reasonable amount of uncertainty about the accuracy of leader effect estimates for individual leaders. This means if the leader effect estimate is modest, we are not able to rule out that it may be zero (and likewise if the least squares leader effect is zero or negative then we can't rule out is positive). A key point of our paper is that leader growth averages should be treated with caution – whether positive or negative.

$$(15) \qquad RMSE = \sqrt{\frac{1}{L}\sum_{i=1..L}\left(\mu_i - \hat{\mu}_i\right)^2} \text{ where } \hat{\mu}_i = \hat{\psi}(\bar{g}_{ic} - \hat{\gamma}\bar{g}_{-ic})$$

In Panel B Table 3 we use estimated variance components using the unbalanced panel adjusted method (SA), rather than the "true" variance components. The estimated values of $\hat{\psi}_{sa}$ and $\hat{\gamma}_{sa}$ are very similar using the unbalanced panel methodology to their analogs using the true SD – for example, with country effects $\hat{\psi}_{sa} = 0.28$ and $\hat{\gamma}_{sa} = 0.78$ vs $\psi = 0.28$ and $\gamma = 0.78$ constructed with true variance components (mean across leaders). SA also performs well in terms of unbiasedness, where estimates of λ are close to one, and within one (Monte Carlo) standard error. The RMSE is only slightly above the one when we know the true variance components. For example, with country effects $RMSE_{sa} = 1.31\%$ vs 1.28% when we know the true variance components.

In Panel C Table 3 we perform the same exercise, but using the standard random effects methodology to calculate the error components. As expected (based on the results in Table 2), the RE method does substantially worse than SA in several dimensions. First, the mean estimates of $\hat{\psi}_{re}$ are around half as large as those using the true variance components (with country effect $\hat{\psi}_{re} = 0.16$, vs 0.28 with the true variance components), and the mean is more volatile across bootstrap replications (though estimates of γ are quite accurate). Second, estimates of λ_{re} are far from one, indicating upward bias. While the estimates of bias are insignificant, the reason for this is the huge standard errors across MC draws, which also indicates a lack of reliability.[30] In terms of the RMSE error, the RE method also does worse than SA, though the difference is not great. For example, with country effects $RMSE_{re} = 1.36\%$, vs $RMSE_{SA} = 1.31\%$, and a value of 1.28% when we know the true variance components.

One of the reasons the RE method does badly is when error variances come from different distributions. In Appendix Table 3, we perform the MC exercise separately for autocracies and democracies, where the error variance is homogeneous within each type. For autocracies, estimates of ψ_{re} are closer to ψ_{true} than ψ_{sa} ($\psi_{re} = \psi_{true} = 0.23$ vs $\psi_{sa} = 0.3$, with country effects), and other estimates ψ_{re} and γ_{re} are just as accurate as ψ_{sa} and γ_{re}. The RE method generally performs slightly worse than SA in terms of bias, and also RMSE, which means that overall we argue that the SA results are more reliable than the RE results.

Using the true variance components --- which we assume for the MC exercise are the same for autocracies and democracies -- generates substantial differences in the reliability and accuracy of leader growth estimates for autocracies and democracies (Appendix Table 3). MC estimates of the reliability factor ψ are smaller in autocracies ($\psi_{aut} = 0.23$) than democracies ($\psi_{dem} = 0.4$) using the true variance components, because the growth process is more noisy in autocracies than democracies (a higher σ_e). The estimate of γ_{aut} is also smaller than γ_{dem}, despite assuming the same σ_c , because the sample for democracies is longer, which allows the model to better distinguish the country effect from a string of good leaders or lucky growth draws (a lower σ_e helps here too). Combined this means that we should put a lot less weight on leader growth averages in autocracies than democracies. Even when we adjust optimally for these factors, there is still more uncertainty around leader quality is autocracies than

[30] The standard errors can be this large because occasionally the RE estimates a very small σ_u , which results in very small leader estimates, and hence very large estimates of λ.

democracies -- the MSE of leader estimates is lower in democracies ($RMSE_{dem} = 1.17\%$) vs autocracies ($RMSE_{aut} = 1.32\%$).

These differences between autocracies and democracies also apply using estimated variances components rather than true variance components. That is, the estimated reliability factors (ψ) and country effects (γ) are smaller under autocracies than democracies, and as the RMSE and bias tends to be smaller under autocracies than democracies (Appendix Table 3).

In sum, our method suggests that even in the presence of a modest true leader SD – such as estimated in Jones and Olken (2005) – the leader growth average is mostly uninformative about the contribution of that leader to growth – especially if the leader is autocrat. To see this, note that the simulated reliability factors in Appendix Table 3 use JO's estimate of the *unobservable* leader SD of 1.5%. The JO findings of strong variation in leader quality are still compatible with low reliability factors because the iid noise can still swamp the variation in μ_i.The reliability/shrinkage factor, for example, about 0.23 for autocrats on average (Appendix Table 3) with a regression of least-squares leader estimates on average growth rates suggesting that a 1ppt increase in the average growth rate leads to a 0.15ppt increase in the least squares leader effect. Hence, even where there is considerable evidence of underlying variation in leader quality, it is hard to infer whether a *particular* leader is of high quality.

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) & (2). The variance components calculated in this section are mostly of interest as a building block for the calculation of the reliability statistic and least squares leader estimates in the next section, though 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. The SD of the leader component (SD(μ_i)) measures the underlying variation in the distribution of leader quality (each leader then draws a quality from this distribution). The standard deviation of the least squares leader estimate (SD($\hat{\mu}_i$)) (to be discussed in the next section) measures the variation in estimated effects on growth of individual leaders in the data (which depends not only on leader quality but also depends on other leader-specific circumstances that affect ability to detect leader quality, like leader tenure).

Table 4: Estimates of Variances Components - SA method (unbalanced panel adjustment)

Sample:	Panel A: All leaders				Panel B: Autocrats				Panel C: Democrats			
Dataset:	PWT 7.1 Data		PWT9 Data		PWT 7.1 Data		PWT9 Data		PWT 7.1 Data		PWT9 Data	
Leader SD	1.33%	0.89%	1.73%	1.42%	1.56%	1.04%	1.96%	1.52%	1.31%	0.96%	1.44%	0.99%
	[0.3%]	[0.38%]	[0.19%]	[0.18%]	[0.29%]	[0.42%]	[0.21%]	[0.2%]	[0.37%]	[0.23%]	[0.18%]	[0.22%]
iid Error	5.75%	5.50%	4.96%	4.73%	6.21%	5.93%	5.39%	5.13%	3.52%	2.67%	2.89%	2.22%
	[0.24%]	[0.24%]	[0.22%]	[0.22%]	[0.26%]	[0.26%]	[0.25%]	[0.24%]	[0.38%]	[0.32%]	[0.26%]	[0.26%]
Country SD	1.45%	1.26%	1.52%	1.39%	1.55%	1.37%	1.64%	1.51%	0.68%	0.39%	0.65%	0.40%
	[0.1%]	[0.1%]	[0.14%]	[0.15%]	[0.14%]	[0.25%]	[0.27%]	[0.3%]	[0.1%]	[0.04%]	[0.12%]	[0.05%]
(p-value country SD=0)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.85	0.23	0.52
Leader variance share	4.78%	2.45%	10.0%	7.64%	5.60%	2.86%	10.79%	7.49%	11.81%	11.24%	19.16%	16.10%
YearXContinent FE	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes
Commodity Pr	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes
Obs / Leaders	6706 /1102		7147 /1155		5214 /783		5531 / 811		1492 /319		1616 /344	

one regresses the per capita real growth rate (by different datasets PWT7.1 or PWT9) on obervables: time by continent FE and/or country-specific commodity price indices (based on that country's exports over 2003-07). Then one collects the residuals and runs an regression of the residuals on country dummies, and decompose the error into a leader effect and a iid error, using the unbalanced panel adjusted random effects estimator (sa in stata). Democracies are countries with an average polity score above 7.5 and autocracies are all other countries. Standard errors are calculated using a country-level block bootstrap with 500 replications. Outliers with greater than 40% growth in a particular year (in absolute value) and Libera dropped.

Using unbalanced panel adjustment method (SA), the unobserved leader component is estimated to have SD of 0.9-1.75% ($\hat{\sigma}_{\mu}$), depending on the specification, regime type and dataset (Table 4). Without controls for observables, the all-leader SD is around 1.3% in PWT 7.1 data, and 1.75% in PWT9. However, our preferred estimates are when we control for observables such as the international business cycle (continent X year FEs), as well as country-specific commodity price indices for commodity exporters.[31] Controlling for these lowers the estimated leader SD to 0.9% and 1.42% in PWT7.1 and PWT 9 respectively. The latter estimate is very similar the 1.47% leader SD estimated in Jones and Olken (2005) (assuming no autocorrelation in leader quality).[32] However, note that these estimates are imprecise – country-level block bootstrap standard errors (in brackets) are usually around 0.2-0.4%.[33]

Using the RE method (Appendix Table 4), estimates of the unobserved leader quality SD ($\hat{\sigma}_{\mu}$) are generally zero for autocrats (and overall). For democracies, however, the leader quality SD is around 0.7-1%, which is far from zero, and on average around 0.2% below that of the unbalanced panel adjustment estimates (SA). So how seriously should we take these results? First, we interpret the zeros as a "small" unobserved leader SD rather than literally zero. The exact zero is a corner solution as leader SD estimate is backed out from the difference between the variance of the leader growth average and the adjusted noise (Equation 10). When the difference is negative, the model reports a zero, even if it is just the case that the leader growth average is small and measured with much noise. Second, the RE method is not always less accurate than the SA method. Monte Carlo estimates of the leader effect (Section 4) favored the SA method overall and in particular when there were heterogeneous iid error variances. However, RE did better for autocrats, where we assumed the same iid error variance across all leaders. In practice of course there is a lot heterogeneity among autocracies – for example year-to-year growth might be less volatile in middle income autocracies than low income countries – and this is a reason to suggest the RE method might perform less well in practice. Nonetheless, it does suggest we interpret the SA results -- which have larger leader effects -- with some caution.

As autocrats are supposed to have greater control over the countries they govern, many in the literature argue that economic growth will be more sensitive to autocratic leader quality (or alternatively autocratic quality varies more) than democratic leader quality (i.e. an autocrat is a "risky bet"). We find only limited evidence of this, and it very much depends on the growth dataset and econometric method used. Estimates of the leader SD are only marginally larger for autocrats than for than for democrats using PWT 7.1, though the gap is larger (around 0.5%) for PWT9. Specifically, the estimated autocratic leader SDs are around 1% for PWT 7.1 and 1.5% for PWT9 in the preferred specification (with controls for observables). The democratic leader SD is around 1% in both cases. With the RE methodology (in the appendix), the democratic leader effects are always larger - and by a significant margin – because the autocratic leader SD is zero. As mentioned above, we interpret these results as a small unobserved autocratic leader SD, and not one that is exactly zero.

---

[31] These are calculated using the Bartik-style approach. We interact the share of commodity exports as proportion of non-resource GDP over 2003-07 from Borensztein et al (2013) with real price variation in the commodity prices from Jacks (2013).

[32] It should be noted that almost all of the reduction in the error is due to the continent X year FEs, with commodity prices making only a marginal contribution. Using Bazzi and Blattman's (2014) commodity price series produces similar results and is reported in the Appendix (controlling for civil wars using the PRIO series also has little effect). Estimated leader effects are slightly larger using only year dummies (rather than continent X year FEs), and are slightly smaller using interactions with subcontinental regions rather than continents (though those sometimes only include a small number of countries) --- reported in the Appendix. Results using the World Bank's World Development Indicators are very similar to those using PWT 9 data, and are reported in the Appendix.

[33] Estimates are quite close to mean of the bootstrap distribution.

As previewed earlier, the year-to-year growth variation (the iid error component) is much more noisy in autocracies than democracies, which is important for identifying the size of the least-squares leader estimate in the next section. In our preferred specification, which controls for observables, the iid error is around of 5-6% in autocracies vs around 2.5% in democracies, with a larger error with PWT7.1 data than PWT9 (4.75-5.5% for all leaders pooled). This is mostly because autocratic countries are poorer on average, and poorer countries tend to experience more noisy growth rates (rather than being determined entirely by the political regime). In the calculations of individual least squares leader estimates in the next section, we estimate the iid errors separately for each country. Country effects are also much larger for autocracies than democracies. The estimated SD of the country effect for autocracies is around 1.4-1.5% for autocracies and statistically significant at the 1% level, whereas for democracies the country effects are around 0.4% and insignificantly different from zero.[34]

The different sizes of iid errors (and country effects) mean that as a share of total unexplained variation in growth, autocratic leaders are much less important than democratic leaders. Using the SA method (Table 3), around 3-7% of variation in growth explained by autocratic leader quality, vs 11-16% for democrats – which suggests that leaders are *relatively* more important for growth in democracies than autocracies. In the RE method (Table 4 in the Appendix), the proportion of variation in growth due to leaders is mechanically much larger for democracies than autocracies as the latter is zero.

Section 6: Least-squares estimates of the leader effects

Now that we have estimates of the variance components (from Table 4), 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). We choose the SA estimates of the variance of the underlying leader quality SD (σ_u) – from which the actual leaders make unobserved draws (autocratic leader estimates are trivially zero using the RE estimates of the leader SD). Our estimates of least squares leader effects are separately on samples of democratic and autocratic countries with the iid SD calculated on a country-by-country basis.[35] Pooled estimate the aggregate the leader estimates calculated separately for autocrats and democrats based on the respective samples.

The least-squares leader effects we estimate in the data reflect how much growth we can attribute to a particular leader. The reliability/shrinkage factor (which is a function of tenure and the variance components), determines whether a high average growth rate is likely reflective of true leader quality (if there is a lot of variation in the unobserved underlying leader quality component), or whether it is likely to reflect more iid noise or country-specific factors.

Across the whole sample of over 1000 leaders, the standard deviation of the estimated least squares leader effect $\hat{\mu}_i$ varies substantially across datasets because the shrinkage/reliability factor depends on the estimates of σ_u from Table 4. Using PWT 7.1 growth data and estimates of σ_u, the standard deviation of the estimated least squares

[34] RE and SA methods estimate almost identical iid errors. They also estimate almost identical country effects. In MC simulations, the country SD tended to be biased upwards, though the tests for significant country effects always correctly identified country effects when they were present, and failed to reject the null hypothesis when they were not.

[35] That is, each step of the algorithm in Section 2 is done separately for autocrats and democrats – the regressions on observables, and calculating true variance components for leaders and country effects (the iid errors are calculated country-by-country). Controlling for international business cycles separately for autocrats and democrats means that in cases where there is only a small number of democratic/autocratic countries in a continent, continent-by-year FE will absorb most variation in growth. Although this sub-optimal, it doesn't have much effect on the results as it only affects a few countries.

leader effect $\hat{\mu}_l$ is around 0.5%, whereas it is around 0.9% using PWT9 (Table 5). In both cases the "shrinkage" is substantial: the standard deviation of the least squares leader effect is around a sixth to a third of the raw leader growth average (which in turn is around 3.25%). This means the average growth rate under a leader is relatively uninformative about which leaders are "good" for growth and which leaders are "bad" for growth, and so the average growth rates are "shrunk" towards zero. As we will see in the next section, this applies particularly strongly for the leaders with the highest and lowest growth averages – as those leaders tend to be from countries with particularly noisy growth processes. The small reliability factors partly reflect the severe noise in average growth rates, especially for shorter tenures. However, even relatively long tenures are surprisingly uninformative for most leaders, because we then have the separate problem of distinguishing the leader effect from the country effect. Under some circumstances – high recorded growth under the leader relative to the country effect and a long tenure -- a strong estimated leader effect does emerge. Even then, regardless of tenure or recorded growth performance, no leader has an estimated least-squares leader effect $\hat{\mu}_l$ of more than around 3.8% in absolute value for PWT 9 or 1.8% for PWT7.1 (both demeaned), and there are very few leaders with a leader effect greater than 1% in absolute value (this applies even more using estimates from the PWT7.1 dataset).[36]

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 (for all leaders with a tenure of at least 3 years using PWT9 data). If the average growth rate under each leader was 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 (correlation of around 0.7), it is much flatter than the diagonal (slope=0.24). The equivalent figure using PWT 7.1 (Appendix Figure 3) looks qualitatively similar, though the slope is flatter at around 0.13.

The path of adjustment from raw leader growth average to least-squares leader estimate is shown in Figure 3 and involves three steps – removing observables, removing country effects and "shrinking" ($\psi \ll 1$) – with the last being the most important for most leaders. Figure 3B shows the effect of removing observables (continent x year FE and commodity prices), 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 (similar for PWT7.1 data in Appendix Figure 3B). This adjustment also makes growth rates mean zero.

The adjustment 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. Figure 3C plots the leader growth average after removing observables (x-axis) vs leader growth average adjusted for country effects and observables on the y-axis.[37] The two are highly correlated, with a slope of 0.93 (0.95 for PWT7.1 growth data in Appendix Figure 3C), which is close to the 45 degree line, suggesting on average the country effect does not dampen much the leader effects. However, the adjustment for country effects is important in improving the fit for particular leaders, where this term ($\hat{\gamma}\overline{g}_{-ic}$) can affect the leader growth average by up to 3-4ppts (though for most leaders it is 1ppt or less). The size of this impact depends on three factors: the size of the country adjustment factor γ, the size of average growth rates under other

---

[36] This refers to a common sample of leaders across PWT9 and PWT7.1. That does remove some leaders from the PWT9 dataset who were in office during 2011-14 . But the majority of the removed leaders are from countries missing from the other PWT dataset, which raises questions about the reliability of those data.

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

leaders and the correlations of these factors with each other and with the leader growth average. Because this adjustment is only weakly related to the size of the leader growth average, it does not systematically reduce the largest (in absolute value) leader effects.

Figure 3: Least Squares Leader Effect Estimates – Pooled Sample

Notes: All figures formed on a common PWT7.1/PWT9 sample. See Appendix Figure 3 for analogous figures A-D using PWT7.1 data.

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 reliability/shrinkage factor ψ. The figure shows the higher *adjusted* average growth rate is only weakly associated with an increase in the least squares leader effect (slope =0.36) due to a small average value of the reliability factor ψ. For PWT 7.1 data, Appendix Figure 3D shows a qualitatively similar relationship, but a line that is much flatter (slope=0.19). For both PWT9 and PWT7.1 the vast majority of difference between the raw leader growth average and LS leader effect in Figure 3A (and Figure Appendix 3A) is due to shrinkage in the final step.

We plot the size of the reliability factor (ψ, y-axis) vs the tenure of the leader (x-axis) in Figure 3F for both PWT9 data and PWT7.1 – as this is where they differ quantitatively. The mean reliability/shrinkage factor in PWT9 data is around 0.42, larger than the mean reliability factor in PWT7.1 data of 0.27 based mostly on the larger estimate of σ_u in Table 3.[38] On average, leaders with a longer tenure have a much higher reliability factor as a longer tenure helps to average out the iid error yielding a correlation between ψ and tenure of 0.25. This is the same relationship as shown in a simple model in Figure 1. However, leaders of longer tenures are usually in poorer countries which have higher iid errors, which is an offsetting factor (each extra year of tenure also adds successively less to the reliability factor). But the dominant feature of Figure 3F is that for leaders of almost any tenure, there is a wide range of reliability factors ranging from almost zero to almost one, and so without an explicit adjustment (as we do in this paper), it is very difficult to know how much information is in the leader growth average.

The other adjustment to the leader growth average is γ (reliability/shrinkage factor for country effects) which is shown in Figure 2E. This also varies substantially across leaders – mostly because of variation in the iid SD - because the growth performance under other leaders can also be attributed to the iid error (rather than the country effect) if the former is large. However, it is also the case that for leaders with a long tenure, there is less information on the performance of *other* leaders, which limits our ability to infer the size of the country effect. That is, the shorter sample of other leaders, the more difficult it is to distinguish between country effects and the talent (or lack thereof) of other leaders. This effect partly offsets the positive relationship between ψ and tenure in Figure 3F.

Democracies vs Autocracies

In the literature there is a consensus that autocrats have a larger effect on growth than democrats, in part because of their greater variability of personalities (tyrant through to benevolent autocrat), and in part because of fewer checks on their power (see literature review in the Introduction). This was already tested in the variance decomposition in Table 4. Least squares leader effects do not measure only variation in quality (the signal) but also our ability to detect that signal due to the amount of noise (reflecting iid error term and leader tenure). As such the variance of the LS effects are not a pure test of the higher autocratic variance hypothesis, but rather a test of the extent that we can observe good and bad autocratic vs democratic leaders. In Table 5 we find mixed evidence (depending on the dataset): after adjusting for iid noise and country effects, there is more variation in estimated least squares leader effects $\hat{\mu}_i$ in democracies than autocracies using the PWT7.1 dataset, though more variation for autocracies in PWT9.

Table 5: Least Squares Leader Estimates

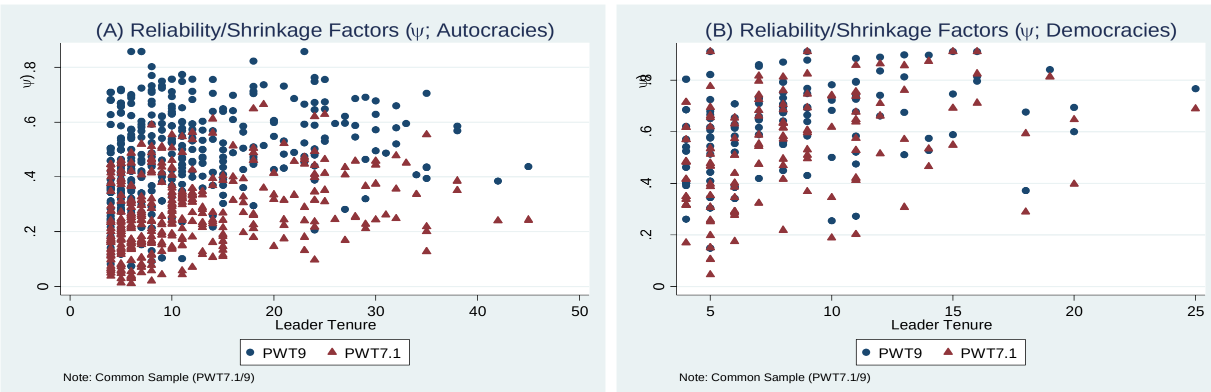
	All leaders (Pooled)		Autocracies		Democracies		Ratio (Aut/Dem)	
	PWT 7.1	PWT 9	PWT 7.1	PWT 9	PWT 7.1	PWT 9	PWT 7.1	PWT 9
SD(average growth under leader)	3.32%	3.18%	3.58%	3.45%	2.55%	2.31%	1.41	1.50
SD(least squares leader effect)	0.54%	0.87%	0.52%	0.95%	0.59%	0.63%	0.88	1.51
Leaders	1069		750		319			

Notes: This table uses two methods to calculate the contribution of leaders to growth. The first method is the average of unadjusted per capita growth rate during the leader's tenure. The second method is the least squares estimates of the leader effect as described in Section 2. We apply each of the methods to either the whole sample of leaders, or only autocrats or only democrats, and calculate the standard deviation of the leader contribution across leaders. Democracies are defined as countries with an average polity score of 7.5 or above (with the rest being autocracies). The population leader SD and country effects are calculated separately for autocracies and democracies, and the iid SD is calculated individually for each country. The pooled sample consisted of the same leader effects as in the autocratic and democratic leader samples, which are then pooled.

[38] In Figure 3F, there are many data points on top of each other for low tenures, which is why the mean is lower than what one might expect from a visual inspection.

Like others in the literature, we find that the average growth rate is more variable under autocrats than democrats. In the first row of Table 5 we see that the standard deviation of the average growth rate under leaders in autocracies is almost 50% higher than that in democracies.[39] However, most of this difference is because of either short-run noise, observable correlates or country-level factors. In the second row of Table 5 we calculate the standard deviation of the least-squares leader effects which suggests the gap between autocracies and democracies is greatly reduced. For PWT7.1 data, the SD of the least squares leader effect across democrats is marginally *larger* across democrats than across autocrats, though for PWT9 it is still larger for autocrats.

Figure 4: Least Squares Leader Effect Estimates – Democracies vs Autocracies



The most important factor for reducing the SD of leader effects is the reliability/shrinkage factor (ψ), which is plotted for autocrats and democrats respectively in Figures 4A and 4B. For a given leader tenure, the reliability factors for democrats are much larger than for autocrats (around 1.3 times for PWT9 and 2 times for PWT7.1). This is mostly because the iid SD is much lower for democracies than autocracies (about half as large). However, it is also the case that there are very few democrats with very long tenures – which have higher reliability/shrinkage factors – whereas longer tenured autocrats are more common.

[39] Note that we remove outliers (Section 2), which strongly affect calculations like this one.

Section 7: Who are the best and worst leaders?

Table 6 shows the best 25 leaders while Table 7 shows the worst 25 leaders as measured by their estimated leader effects using PWT9 data (these are plotted in Figure 4 panels A-B on the y axis). These rankings reflect not only leader quality but also our ability to detect that quality. We will find a strong leader effect when the growth rate signal is strong relative to the noise. A strong signal means a high residual growth rate after the raw growth rate is adjusted for commodity prices and region-year effects, and then compared to all other leaders' such residuals for the same country. Noise tends to shrink all of the leader estimates but noise is less of a factor for some leaders than others. Low noise means a low country-varying Standard Deviation of the iid error term, and then possibly a long tenure to average out this noise. Reviewing some of the best and worst leader effects helps show why the leader estimates diverge so much from the prevailing practice of giving the leader credit for all of the raw average growth average during his tenure.

Table 6: Best 25 leaders according to PWT9 least squares estimates (tenure 4yrs+, country-specific iid SD)

		Leader Rank (PWT9)		Leader Rank (PWT7.1)		LS Leader Effect		Growth Average		Demo-	
Leader name	**Country**	**Least Sq**	**Growth Ave**	**Least Sq**	**Growth Ave**	**PWT9**	**PWT 7.1**	**PWT9**	**PWT 7.1**	**cracy**	**Tenure**
Chun Doo Hwan	KOR	1	6	14	15	3.3%	1.4%	8.3%	7.7%	0	8
Hun Sen	KHM	2		4		2.8%	1.8%	5.2%	5.8%	0	25
Khama	BWA	3	1	11	1	2.8%	1.5%	9.6%	10.2%	0	15
Medici	BRA	4	8	3	8	2.6%	1.8%	8.1%	8.4%	0	5
Kubitschek	BRA	5		6		2.6%	1.6%	5.9%	5.9%	0	5
Razak	MYS	6	15	19	10	2.5%	1.2%	7.7%	8.1%	0	6
Museveni	UGA	7		8		2.5%	1.5%	3.4%	3.1%	0	28
Lee Kuan Yew	SGP	8				2.4%	0.9%	6.3%	5.6%	0	30
Chiang Ching-Kuo	TWN	9				2.4%	0.9%	6.4%	5.7%	0	10
Manmohan Singh	IND	10		7	21	2.3%	1.6%	6.0%	6.8%	1	10
Soares	PRT	11		5		2.3%	1.6%	3.3%	3.7%	0	10
Zia	PAK	12		25		2.3%	1.1%	3.3%	3.3%	0	11
Cristiani	SLV	13		10		2.2%	1.5%	4.1%	3.4%	0	5
Elbegdorj	MNG	14	3			2.2%	0.1%	8.7%	7.0%	0	5
Nkurunziza	BDI	15	17			2.1%	-0.1%	7.5%	1.2%	0	9
Chissano	MOZ	16		9		2.0%	1.5%	4.3%	4.1%	0	19
Berisha	ALB	17				2.0%	1.0%	5.3%	5.9%	0	13
Hasina Wazed	BGD	18				1.9%	1.0%	4.0%	3.5%	0	10
Paul Kagame	RWA	19				1.9%	0.3%	6.1%	3.4%	0	20
Rodriguez Lara	ECU	20	11		11	1.8%	0.8%	8.0%	8.0%	0	4
Raab	AUT	21		1		1.8%	1.8%	6.1%	6.2%	1	8
Rajapakse	LKA	22				1.8%	0.2%	6.0%	4.9%	0	9
Grunitzky	TGO	23	16		13	1.7%	1.0%	7.6%	7.9%	0	4
Reagan	USA	24		2		1.7%	1.8%	2.6%	2.6%	1	8
Lopez Portillo	MEX	25				1.7%	0.6%	3.5%	3.4%	0	6

Notes: This table ranks leaders according to their least squares leader effect using PWT9 growth data. The least squares leader effect starts with the average per capita growth rate residual (after removing the international business cycle and effects of commodity prices), then removes estimated country effects, and the remaining average is "shrunk" towards to zero depending on the signal to noise ratio of the underlying growth rate. This is the LS leader effect. "Growth average" in the table is the unadjusted average growth rate of per capita GDp during the leaders' tenure. Democracies are countries with an average polity score above 7.5. The PWT 7.1 series are calculated in the same way, but using PWT7.1 data. Blank indicates rank outside the top 25. Tenure to 2014.

The region-year effects 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 leader average growth rates. This adjustment will be more or less depending on the timing of commodity prices and regional business cycles for each leader. So for example, Chun Doo Hwan of South Korea at the top of the rankings show little such adjustment (less than 1 percentage point), while Lee Kwan Yew is further down the rankings because of a larger adjustment (more than 2 percentage points). Asia was doing better during Lee Kwan Yew's term than during Chun Doo Hwan's, and so Lee's performance is discounted more than Chun Doo Hwan's.

The residual after adjustment will then be compared to the same residual for other leaders for the same country. Some of the most 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, Yoweri Museveni of Uganda, Pierre Nkurunziza of Burundi, Joaquim Chissano of Mozambique, and Paul Kagame of Rwanda, and to a lesser extent Manmohan Singh of India. The flip side of this will be that some of those disastrous leaders will show up among the worst leaders in Table 7: Pol Pot (#11 worst) and Lon Nol (#2 worst) of Cambodia, Idi Amin of Uganda (#1 worst), and Samora Machel of Mozambique (#12 worst).

Table 7: Worst 25 leaders according to PWT9 least squares estimates (tenure 4yrs+, country-specific iid SD)- Rank from Last

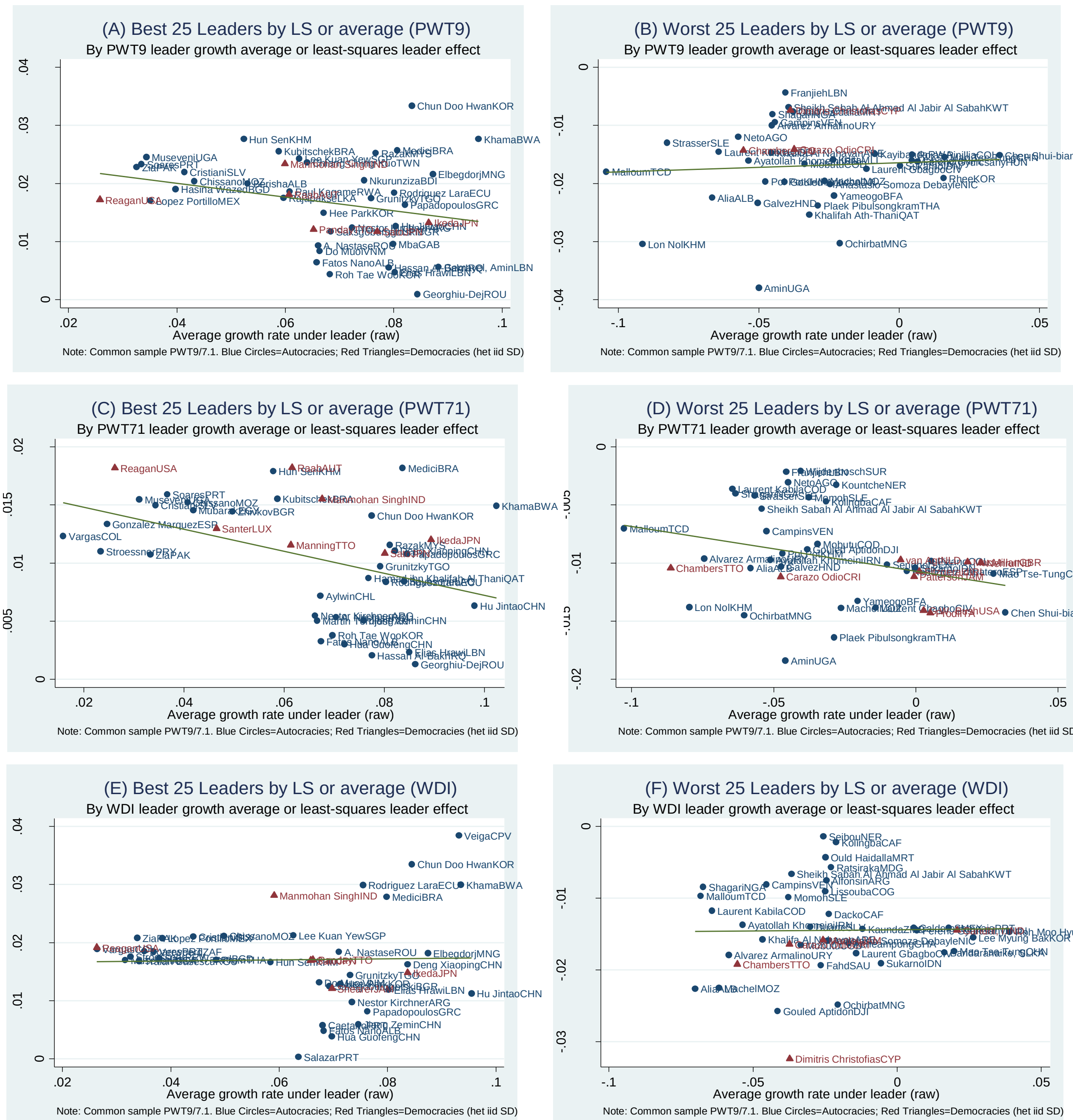
		Leader Rank (PWT9)		Leader Rank (PWT7.1)		LS Leader Effect		Growth Average		Demo-	
Leader name	**Country**	**Least Sq**	**Growth Ave**	**Least Sq**	**Growth Ave**	**PWT9**	**PWT 7.1**	**PWT9**	**PWT 7.1**	**cracy**	**Tenure**
Amin	UGA	1	10	1	16	-3.8%	-1.8%	-5.0%	-4.6%	0	8
Lon Nol	KHM	2	2	9	3	-3.0%	-1.4%	-9.1%	-8.0%	0	5
Ochirbat	MNG	3		3	7	-3.0%	-1.5%	-2.1%	-6.0%	0	7
Khalifah Ath-Thani	QAT	4	23			-2.5%	-0.5%	-3.2%	-0.4%	0	23
Plaek Pibulsongkram	THA	5	24	2	24	-2.4%	-1.6%	-2.9%	-2.9%	0	7
Galvez	HND	6	9	19	14	-2.3%	-1.0%	-5.0%	-4.7%	0	4
Alia	ALB	7	4	17	8	-2.2%	-1.0%	-6.7%	-5.8%	0	7
Yameogo	BFA	8		10		-2.2%	-1.3%	-2.3%	-2.0%	0	7
Anastasio Somoza Del	NIC	9				-2.0%	-0.5%	-2.5%	0.6%	0	12
Gouled Aptidon	DJI	10	16		20	-2.0%	-0.9%	-4.1%	-3.8%	0	23
Pol Pot	KHM	11	11		15	-2.0%	-0.9%	-4.8%	-4.7%	0	4
Machel	MOZ	12	25	7		-2.0%	-1.4%	-2.6%	-2.6%	0	12
Rhee	KOR	13				-1.9%	-0.9%	1.6%	1.6%	0	7
Malloum	TCD	14	1		1	-1.8%	-0.7%	-10.4%	-10.3%	0	4
Laurent Gbagbo	CIV	15		8		-1.7%	-1.4%	-1.2%	-1.4%	0	11
Sukarno	IDN	16		18		-1.7%	-1.0%	0.0%	0.0%	0	6
Mobutu	COD	17	22		22	-1.7%	-0.8%	-3.4%	-3.5%	0	32
Ferenc Gyurcsany	HUN	18				-1.6%	-0.9%	0.7%	0.6%	0	5
Ayatollah Khomeini	IRN	19	8	25	12	-1.6%	-1.0%	-5.4%	-5.1%	0	10
Mahendra	NPL	20				-1.6%	-0.9%	0.5%	0.5%	0	12
Keita	MLI	21				-1.6%	-0.8%	-2.4%	-2.4%	0	8
Mao Tse-Tung	CHN	22		13		-1.5%	-1.1%	1.6%	2.7%	0	24
Chen Shui-bian	TWN	23		5		-1.5%	-1.4%	3.6%	3.2%	0	8
Rojas Pinillia	COL	24				-1.5%	-0.6%	0.5%	0.8%	0	4
Kayibanda	RWA	25				-1.5%	-0.4%	-0.9%	-0.3%	0	13

Notes: This table ranks leaders according to their least squares leader effect using PWT9 growth data, starting from the most negative growth effect and working backwards. The least squares leader effect starts with the average per capita growth rate residual (after removing the international business cycle and effects of commodity prices), then removes estimated country effects, and the remaining average is "shrunk" towards to zero depending on the signal to noise ratio of the underlying growth rate. This is the LS leader effect. "Growth average" in the table is the unadjusted average growth rate of per capita GDp during the leaders' tenure. Democracies are countries with an average polity score above 7.5. The PWT 7.1 series are calculated in the same way, but using PWT7.1 data. "Rank" is from last of 635 leaders with tenure of at least 4 years (and 4 years of PWT9 data). Tenure to 2014. Blank indicates rank outside bottom 25.

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, Khama, and Chiang Ching-Kuo of Taiwan. Similarly, unsurprising bad leaders are those who have unusually negative growth even for a low growth country: such as Mobutu, the Ayatollah Khomeini, and Anastasio Somoza Debayle of Nicaragua, or a bad growth rate for a high growth country such as Mao Tse-Tung for China and Sukarno for Indonesia.

Finally, detecting the leader effect means relatively low noise. This happens with some of the top leaders due to long tenures averaging out noise (approximate tenures in brackets): Hun Sen (25 years), Khama (15), Museveni (28), Lee Kuan Yew (30), and Kagame (20+). Similarly, some of the worst leader effects also reflect long tenures: Mao (24), Mobutu (32), Machel (12), and Somoza (12). 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.

Figure 5: Best and worst leaders by different datasets



Other top leader estimates have shorter tenures 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.[40] Low country noise is particularly important in detecting effects of democratic leaders who tend to have shorter tenures than autocratic leaders. So for example, Ronald Reagan (#24) has a detectable effect because of very low country noise in the US. Another democratic leader with a strong leader effect, Manmohan Singh (#10) of India, also had relatively low country noise compared to other low income countries.

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 (#85) of China, whose leader effect suffered from the high growth of most other Chinese leaders and from very high noise in China (standard deviation of the error term more than 6 percentage points). Another famous leader missing is Chiang Kai-Shek in Taiwan (#144), who had a large downward adjustment because of commodity prices and regional business cycles and then was no better at producing growth than other Taiwanese leaders. Conversely, Table 6 (for the top leader) features many names that have received little attention as growth stars.

The good news for the benevolent autocrat hypothesis is that we do confirm strong estimated leader effects for some celebrated leaders, such as Khama, Lee Kuan Yew, Museveni, and Kagame. The news that is not as good is that other celebrated leaders fail to show a strong leader effect, such as Deng Xiaoping and Chiang Kai-Shek, while other uncelebrated leaders show up instead in the top ranks. The same thing occurs among the worst leaders table, where famous disasters are mixed with many little known leaders. We conjecture that this is because identification of benevolent and malevolent autocrats is based on raw growth averages of leaders, while we have seen that estimating the leader effect requires major adjustments to that raw growth average.

Robustness across datasets

The next check that we do on our results is to see whether the leader estimates from PWT9 are robust to using our other growth datasets: PWT7.1 and World Bank WDI. The results are mixed: around half (13) of the top 25 leaders from PWT9 are also in the top 25 in PWT7.1 and WDI.[41] Leaders discussed above that are confirmed across datasets include Chun Doo Hwan, Hun Sen, Khama, Museveni, Chissano, and Reagan. Leaders that fail to be validated by other datasets include Lee Kuan Yew, Chiang Ching-Kuo, Pierre Nkurunziza, and Paul Kagame.

Among the worst 25 leaders by leader effect in PWT9, around half (11) of the worst 25 also show up among the worst in the other datasets. Those confirmed across datasets include Idi Amin, Lon Nol, Machel, Sukarno, and Mao. Those that fail to be confirmed as among the 25 worst in other datasets are Pol Pot and Mobutu.[42]

The other datasets also identify different stars based on least squares leader estimates. Leader estimates of autocrats are shrunk more in PWT7.1 than in PWT9, because the variance decomposition for PWT7.1 compared to PWT9 shows a lower underlying SD of autocratic leader quality and higher SD of iid errors in autocracies (from Table 3). There is not much of a change in shrinkage of democratic leader effects from PWT9 to PWT7.1, in part because data in democracies tends to be of better quality (and the size of the leader effect σ_u^2 is almost identical in PWT9 and PWT 7.1). As a result, democratic leaders show up more in the best and worst in PWT7.1 compared to PWT9.

---

[40] 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 could be too simplistic.

[41] Two of these are confirmed by PWT7.1 but not available in WDI because of sample restrictions (omitting the 1950s and omitting Taiwan).

[42] Five of these are validated by PWT7.1 but not available in WDI.

In PWT7.1 (Appendix Table 6), the top two leaders are democrats: Reagan is #2, and #1 is Julius Raab (1954-61) of Austria. Manhoman Singh moves up from #10 in PWT9 to #7 in PWT7.1, and four other democratic leaders move into the top 25 in PWT7.1. Nine new democratic leaders that were not in the PWT9 25 worst now show up among the worst leaders in PWT7.1, including Jawaharlal Nehru and George W. Bush (recall GW Bush is allocated growth during 2009 as he is leader on 1 January of that year).

WDI is notable for showing a few jumps for some leaders (Appendix Table 8 and 9): Hun Sen falls from #2 in PWT9 to #25 in WDI, while Reagan jumps from #24 to #12. Lee Kuan Yew is back among the top in WDI after missing out on PWT7.1. Manmohan Singh is at #5 in WDI, and he and Emilio Medici of Brazil are the only leaders in the top 10 in all 3 datasets. As with PWT7.1, democratic leaders are somewhat more likely to be among the 25 worst leader effects in WDI than occurred with PWT9. Five democratic leaders make the bottom 25, the most well-known being Indira Gandhi.

Another concern about data quality is shown by four leaders of the worst 25 from PWT9 that are missing WDI values, including Idi Amin, Lon Nol, and Pol Pot. Apparently WDI considered the growth data of insufficient quality to report numbers for these cases.[43] We prefer the PWT9 dataset on the basis of methodology and the largest sample, so the confirmation failures on some of the best and worst leaders could just be that the other datasets are worse measures of growth than PWT9, or had reduces coverage of countries and years.

We note also that some of the disastrous leader effects are associated with wars or genocides (such as Burundi, Democratic Republic of the Congo, Rwanda, and Cambodia). It is not surprising that growth datasets disagree on the extreme growth outcomes during and after wars (with WDI sometimes failing to report data at all). These volatile growth estimates during wars affect both the estimates of the worst leaders and the best (by comparison with the worst). Countries with wars also have much higher and more volatile estimates of the standard deviations of the iid error term, which we have seen tends to diminish leader estimate magnitudes in absolute value. (For example, the SD of the iid error in PWT7.1 in Rwanda is 9.1% compared to 6.3% in PWT9, which helps eliminate Kagame as a top leader in PWT7.1).[44] So lack of robustness across datasets for leaders in countries with major wars, like Kagame, Mobutu, Pol Pot, and Nkurunziza is not very surprising. At the same time, as we have seen, discussions sometimes identify leaders like Kagame as confidently good for growth anyway, so there should be more awareness of the unreliability of the growth data around wars and postwar recoveries.

Even acknowledging these caveats, the instability of leader estimates across datasets is a reminder once again of how high is the noise in the growth data, making it difficult to extract a reliable signal on leader effects for any one particular leader. The variance decomposition exercise had already partly confirmed that "leaders matter," but the exercise in this section shows the difficulty of identifying *which* leaders matter.

Comparing best (or worst) leaders by least squares vs leader growth averages

The discussion above has focused on famous instances of best and worst leaders, who have a high or low average growth rate as well as a narrative that capture the attention of policymakers or the public. In the scatter plots in Figure 4 we plot the leader growth average (x axis) against the least squares leader effect (y axis) for best and worst

---

[43] Growth data for WDI starts in 1983 for Uganda and in 1994 for Cambodia. There are also cases where PWT7.1 is missing when PWT9 is observed (and vice versa), but for this exercise we dropped all observations that appear only in ONE of these two datasets.

[44] Volatile growth data also interacts with the ±40% cutoff for an outlier – in 1995, the post-civil war rebound growth in Rwanda is counted as an outlier in PWT7.1, but not in PWT9.

leaders (by either leader growth average or LS leader effect) from the different datasets. The first thing to notice is that among these good or bad leaders, the leader growth average either is unrelated to the least squares leader effect (a zero correlation ("useless") Panels B, E, F), or even worse, there is a negative relationship between average growth under the leader and least squares leader effect ("worse than useless", Panels A, C, D). This is a remarkable result given that that the least squares leader effect is constructed out of the average leader growth rate.[45]

The main reason for the negative slope (or absence of a positive slope) is that those countries which tend to have very high (or very low) growth rates *also* tend to have very volatile growth rates, and for this reason these extreme growth rates are shrunk towards zero. For example, three of the leaders with highest or lowest growth averages come from Lebanon, where high and low growth rates were generated by volatility due to the Lebanese civil war (1975-90). This growth volatility leads to a smaller reliability factor, shrinking the high or lower leader growth average towards zero.

A negative slope can also be generated by country effects (performance under other leaders). The best leaders in the bottom RHS corner of the plots in Figure 5 can have their least squares leader effect reduced because of strong growth under other leaders in the same country (a number of leaders in China fall into this category). On the diagonally opposite side of the same plot are leaders that don't have a spectacular leader growth average themselves but are boosted by poor growth under other leaders in the same country (examples include Hun Sen of Cambodia and a number of others discussed above).

The bottom line is that our methodology produces very different estimates of good and bad leader effects than would be inferred from raw growth averages during the leaders' time in office. The good news for the benevolent autocrat hypothesis is that we do confirm major effects of some celebrated autocrats on growth (given our simple assumptions). The bad news is that different datasets often disagree on the identity of the top benevolent autocrats.

Section 8: Conclusions

In this paper, we presented a method for estimating every leader's impact on growth based on the teacher value added literature. We believe this method is a major improvement over the current practice of giving a leader full credit for the average growth rate during his tenure.

Our main insight from this method is that even under a model where leaders do affect economic growth, and even when we confirm that leaders matter in our variance decomposition, the average growth rate during the tenure of a leader is still mostly uninformative about that particular leader's true contribution to growth. This is most of all because year-to-year growth rates are so volatile, and so the average growth rate under most leaders has lots of noise, and not much signal. Secondarily, plausible controls for commodity prices, business cycles by continent, and country effects reduce further leader effects and make them even less correlated with raw leader growth averages. Moreover, the average growth rates during a leader's tenure vary substantially across growth datasets, so the set of high (and low) growth leaders changes depending on the dataset used.

Our results are mixed on the received wisdom in the literature that the contributions of autocratic leaders to growth are larger than those for democratic leaders. We do find some supportive evidence, though it turns out not to be robust across methods and datasets. In our preferred dataset (PWT9) and method (SA) for the variance

[45] Across all leaders, the leader growth average and least squares leader effect have positive correlation of 0.7, which one can see in Figure 3A. The two measures are positively correlated (Figure 3A, correlation of around 0.7).

decomposition of unobserved components, we actually replicate the Jones and Olken (2005) results of leader quality having a standard deviation of about 1.5 percentage points, and show that this is larger for autocrats than democrats. Our least squares estimates of leader effects tend to be larger in absolute value for leaders with longer tenures (as this helps to even out short-run noise), and long tenures are more likely among autocrats. So our preferred estimates of the 25 best and worst leaders heavily feature autocrats, and we do confirm major growth effects of some celebrated cases of good and bad dictators.

However, these results are not quite as strong as they seem for the hypothesis of autocratic leaders mattering more than democrats. On the variance decomposition, we also find a much larger standard deviation of the iid error term in autocracies compared to democracies. Since we do find that leaders matter in democracies (unlike JO) with a leader SD of about 1 percentage point, it turns out that democratic leader effects explain more of the total variation of growth than do autocratic leader effects. The problem of high autocratic noise shows up in other ways. Some celebrated dictators fail to show up anywhere near the top LS estimates of leader effects (most surprisingly Deng Xiaoping), while other previously unnoticed dictators do show up in top and bottom 25 leaders. Moreover, the results of the previous paragraph are not very robust across methods and datasets. The RE method for the variance decomposition suggests democratic leaders in PWT9 have a larger SD of unobserved quality than do autocrats (even generating a corner solution of zero SD for autocrats). Switching datasets from PWT9 to PWT7.1 generates a similar SD for democratic leaders in the variance decomposition in the SA method and it features many more democratic leaders in the top and bottom 25 least squares leader effect estimates, and PWT7.1 fails to confirm some of the specific best and worst autocrats from PWT9 (Lee Kuan Yew drops out for example). The World Bank WDI growth dataset also shows yet further disagreement on best and worst leaders.

In sum, in a world of tremendous volatility of growth rates, we generally confirm some leaders do matter for growth -- but have trouble discerning which ones they are, including whether they are more likely to be autocrats or democrats.

What are the implications of these results? Most importantly, policymakers and commentators should be much more careful about attributing economic growth to leadership. This is especially true when the tenures of leaders are short and in autocracies, where annual growth is more volatile -- and of course policymakers are usually not going to wait 20 years to pass judgement. And then, as the bias due to the year-to-year growth volatility dies away, there then arises a new difficulty in distinguishing between a long-serving leader effect and a country effect. And even longer-term leader growth averages also need to filter out variables that affect growth but are beyond the leader's control -- commodity prices and regional business cycles (especially the latter). Policymakers should also take into account how much different growth datasets agree or disagree on performance under particular leaders.

All of this creates a great deal of uncertainty on the claim of any one particular autocrat to be "benevolent" for purposes of attracting aid or other foreign or domestic support.

Our results make us wonder whether the theory of benevolent autocrats was too simple. Even seemingly unconstrained autocratic leaders might find it difficult to exert control over the growth rate of the economy. Even if there were a "benevolent" autocrat determined to raise growth, he or she has to solve difficult principal-agent problems to get his/her growth-promoting orders carried out all the way down the government bureaucracy. The autocrat also has to solve a serious knowledge problem getting accurate information on what are the most serious obstacles to growth and/or what are the biggest opportunities for government actions to raise growth. Autocratic leaders also may face many constraints even though they don't face democratic ones, as there are other power centers in autocratic systems that may be able to veto actions contrary to their interests. In sum, the theory of benevolent autocrats producing growth miracles requires strong assumptions about the autocrats' ability to motivate the government bureaucracy, solve knowledge problems, and overcome other elite interests running contrary to growth. And another big assumption was that an autocrat selected through a ruthless process of amassing power would indeed turn out to be benevolent very often.

Development policy could be one of the last refuges of the "Great Man" theory of history, which has been discarded in history itself and in most other social science analysis. We indeed find that even in a model and empirical methodology seemingly favorable to leader effects, most of what is going on in economic growth has little to do with the national leader.

References:

Acemoglu, D., Johnson, S., Robinson, J., Thaicharoen, Y. (2003). "Institutional causes, macroeconomic symptoms: volatility, crises and growth". *Journal of Monetary Economics*. 50: 49–123

Ashworth S., E. Bueno de Mesquita, and A. Friedenberg, (2016) "Learning about Voter Rationality", mimeo, July 12, 2016. http://home.uchicago.edu/~sashwort/shocks.pdf

Almeida H. and D. Ferreira (2002), "Democracy and the Variability of Economic Performance" *Economics and Politics*, 14(3), 225-257.

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

Becker G. (2010). *Democracy or Autocracy: Which is better for Economic Growth?*, Becker-Posner Blog Post (10/10/2010) [link]

Besley, T. and Masa Kudamatsu (2009), "Making Autocracy Work", in Elhanan Helpman (ed), *Institutions and Economic Performance*, Cambridge: Harvard University Press.

Besley, T., J. Montalvo and M. Reynal-Querol (2011), "Do Educated Leaders Matter?", *The Economic Journal*, 121 (August), 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.

Birdsall N. and F. Fukuyama (2011) "The Post-Washington Consensus: Development After the Crisis." *Foreign Affairs* 90, no. 2 (March/April 2011): 51. [link]

Borensztein, E., O. Jeanne, and D. Sandri (2013) "Macro-hedging for commodity exporters", *Journal of Development Economics* 101,105-116

Brady D. and M. Spence (2010) (eds), *Leadership and Growth*, World Bank Publications, Washington DC

Bueno de Mesquita, B. and A. Smith (2011), *The Dictator's Handbook: Why Bad Behavior is Almost Always Good Politics*, PublicAffairs, New York, NY

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.

Commission on Growth and Development (2008), *The Growth Report: Strategies for Sustained Growth and Inclusive Development*, World Bank Publications (6507), Washington DC

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

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

Hayes, R. C., Imai, M. and Shelton, C. A. (2015), "Attribution Error in Economic Voting: Evidence from Trade Shocks." *Economic Inquiry*, 53: 258–275.

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.

Jacks, D. (2013) “From Boom to Bust: A Typology of Real Commodity Prices in the Long Run” NBER WP 18874

Johnson S., W. Larson, C Papageorgiou and A. Subramanian (2013) “Is newer better? Penn World Table Revisions and their impact” *Journal of Monetary Economics* 60, 255 –274

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.

Jones, B and B Olken (2009) “Hit or Miss? The effect of Assassinations on Institutions and War” *American Economic Journal: Macroeconomics* 1 (2): 55-87.

Jones, C. and P. Romer (2010) “The New Kaldor Facts: Ideas, Institutions, Populations and Human Capital”, *American Economic Journal: Macroeconomics 2(1): 224-245*

Kaldor, N. (1961). “Capital Accumulation and Economic Growth.” In The Theory of Capital, ed. F. A. Lutz and D. C. Hague, pp 177–222. New York: St. Martins Press.

Kane, T.., and D. Staiger, (2008) “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation,” NBER working paper 14607

Research Working Paper No. 14607, 2008.

Kinsler, J. (2012) “Assessing Rothstein’s Critique of teacher value-added models” *Quantitative Economics* 3, 333-362

Mobarak, A. (2005) “Democracy, Volatility, and Economic Development”, *Review of Econ. and Statistics*, 87(2): 348-361

Magee C. and J. Doce (2014) “Reconsidering Regime Type and Growth: Lies, Dictatorships, and Statistics” *International Studies Quarterly* (2014) 1–15

Meyersson E. (2014) “Political Man on Horseback: Military Coups and Development” Mimeo Stockholm School of Economics

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

Popa, M. (2013) “Political Leadership Time Horizons and Economic Growth”, Working Paper, Yale University

Quinn D. and J Woolley “Democracy and National Economic Performance: The Preference for Stability” *American Journal of Political Science* 45:634-657

Rodrik, D. (2000), "Institutions for High-Quality Growth: What They Are and How to Acquire Them," NBER Working Paper No. 7540

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

Sah, R. (1991), “Fallibility in human organizations and political systems”. *Journal of Economic Perspectives*. 5: 67–88.

Weede, E., (1996), “Political regime type and variation in economic growth rates” *Constitutional Political Economy* 7, 167–176.

Justin Wolfers, Are Voters Rational? Evidence from Gubernatorial Elections, NBER Working Paper, 2007. http://users.nber.org/~jwolfers/papers/Voterrationality(latest).pdf.

World Bank (2012). *Global Development at a Pivotal Time: A Conversation with World Bank President Dr. Jim Yong Kim*, Speech at the speech at Brookings Institution, July 19, 2012, [link]

World Bank (2013), *Country assistance strategy for the Democratic Republic of the Congo*, Report No. 66158-ZR, April 12, 2013. [link]

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

Yang, B. (2008). “Does democracy lower growth volatility? A dynamic panel analysis”, *Journal of Macroeconomics* 30 (2008) 562–574

Appendix 1: Additional Figures and Tables

Appendix Table 1: Country List

Country Name	Code	Average Polity	Democracy	Country Name	Code	Average Polity	Democracy
Afghanistan	AFG	-5.2	0	Nicaragua	NIC	-0.7	0
Albania	ALB	-0.3	0	Niger	NER	-2.1	0
Algeria	DZA	-5.1	0	Nigeria	NGA	-0.5	0
Angola	AGO	-4.1	0	Oman	OMN	-9.2	0
Argentina	ARG	1.3	0	Pakistan	PAK	1.2	0
Bahrain	BHR	-8.9	0	Panama	PAN	2.2	0
Bangladesh	BGD	0.8	0	Papua New Guinea	PNG	4.1	0
Benin	BEN	-0.2	0	Paraguay	PRY	-2.0	0
Bhutan	BTN	-7.5	0	Peru	PER	3.0	0
Bolivia	BOL	2.0	0	Philippines	PHL	3.2	0
Botswana	BWA	6.9	0	Poland	POL	2.5	0
Brazil	BRA	2.8	0	Portugal	PRT	2.8	0
Bulgaria	BGR	1.8	0	Qatar	QAT	-10.0	0
Burkina Faso	BFA	-3.8	0	Republic of Korea	KOR	1.5	0
Burundi	BDI	-2.6	0	Romania	ROU	-0.3	0
Cabo Verde	CPV	4.1	0	Rwanda	RWA	-5.4	0
Cambodia	KHM	-0.2	0	Saudi Arabia	SAU	-10.0	0
Cameroon	CMR	-6.0	0	Senegal	SEN	-0.4	0
Central African Republic	CAF	-3.4	0	Sierra Leone	SLE	-0.9	0
Chad	TCD	-4.7	0	Singapore	SGP	-1.7	0
Chile	CHL	3.7	0	Solomon Islands	SLB	7.0	0
China	CHN	-7.4	0	Somalia	SOM	-3.2	0
Colombia	COL	6.3	0	South Africa	ZAF	5.8	0
Comoros	COM	1.4	0	Spain	ESP	3.1	0
Congo	COG	-4.8	0	Sri Lanka	LKA	6.0	0
Cuba	CUB	-7.0	0	Sudan (Former)	SDN	-4.0	0
Cote d'Ivoire	CIV	-5.5	0	Suriname	SUR	2.8	0
D.R. of the Congo	COD	-3.6	0	Swaziland	SWZ	-9.0	0
Djibouti	DJI	-3.4	0	Syrian Arab Republic	SYR	-8.0	0
Dominican Republic	DOM	2.4	0	Taiwan	TWN	-1.0	0
Ecuador	ECU	4.0	0	Thailand	THA	1.0	0
Egypt	EGY	-5.8	0	Togo	TGO	-4.9	0
El Salvador	SLV	2.1	0	Tunisia	TUN	-5.5	0
Equatorial Guinea	GNQ	-6.3	0	Turkey	TUR	6.5	0
Ethiopia	ETH	-7.6	0	U.R. of Tanzania: Main	TZA	-4.1	0
Fiji	FJI	4.4	0	Uganda	UGA	-2.8	0
Gabon	GAB	-5.7	0	United Arab Emirates	ARE	-8.0	0
Gambia	GMB	2.3	0	Uruguay	URY	5.6	0
Ghana	GHA	-1.1	0	Venezuela (Bolivarian I	VEN	5.7	0
Greece	GRC	6.2	0	Viet Nam	VNM	-7.0	0
Guatemala	GTM	1.4	0	Zambia	ZMB	-0.9	0
Guinea	GIN	-5.3	0	Zimbabwe	ZWE	-0.9	0
Guinea-Bissau	GNB	-1.4	0	Australia	AUS	10.0	1
Guyana	GUY	1.4	0	Austria	AUT	10.0	1
Haiti	HTI	-4.1	0	Belgium	BEL	9.8	1
Honduras	HND	2.8	0	Canada	CAN	10.0	1
Hungary	HUN	3.0	0	Costa Rica	CRI	10.0	1
Indonesia	IDN	-2.5	0	Cyprus	CYP	8.7	1
Iran (Islamic Republic of)	IRN	-6.4	0	Denmark	DNK	10.0	1
Iraq	IRQ	-6.9	0	Finland	FIN	10.0	1
Jordan	JOR	-6.1	0	France	FRA	8.2	1
Kenya	KEN	-1.6	0	Germany	DEU	10.0	1
Kuwait	KWT	-8.0	0	India	IND	8.7	1
Lao People's DR	LAO	-6.5	0	Ireland	IRL	10.0	1
Lebanon	LBN	2.8	0	Israel	ISR	7.7	1
Lesotho	LSO	0.5	0	Italy	ITA	10.0	1
Liberia	LBR	-2.1	0	Jamaica	JAM	9.6	1
Madagascar	MDG	0.6	0	Japan	JPN	10.0	1
Malawi	MWI	-2.9	0	Luxembourg	LUX	10.0	1
Malaysia	MYS	5.1	0	Mauritius	MUS	9.7	1
Mali	MLI	-1.2	0	Netherlands	NLD	10.0	1
Mauritania	MRT	-5.8	0	New Zealand	NZL	10.0	1
Mexico	MEX	-0.6	0	Norway	NOR	10.0	1
Mongolia	MNG	2.2	0	Sweden	SWE	10.0	1
Morocco	MAR	-6.8	0	Switzerland	CHE	10.0	1
Mozambique	MOZ	-0.9	0	Trinidad and Tobago	TTO	8.9	1
Myanmar	MMR	-6.4	0	United Kingdom	GBR	10.0	1
Nepal	NPL	-2.0	0	United States	USA	10.0	1

Notes: Countries used in the paper. Democracy is defined as an average polity score >7.5. Some countries might be dropped due to outliers (such as Liberia), the fact that they only have one leader for some datasets (Oman), or be missing data for particular datasets (Afghanistan for PWT9)

Appendix Table 1a: Outlier List					
Country Name	**Year**	**Country Code**	**Outliers Dropped**		
			PWT9	**PWT7.1**	**WDI**
Afghanistan	1993	AFG		1	
	2002	AFG		1	
Central African Rep	2013	CAF	1		1
Algeria	1962	DZA	1	1	0
Equatorial Guinea	1996	GNQ	0	1	1
	1997	GNQ	1	1	1
	2001	GNQ	1	0	1
Iran	1970	IRN	1	0	0
Iraq	1990	IRQ	0	0	1
	1991	IRQ	1	1	1
	2003	IRQ	1	1	1
	2004	IRQ	1	1	1
Kuwait	1990	KWT	Dropped due to Iraq occupation		
	1991	KWT			
	1992	KWT	1	0	
Lebanon	1976	LBN	1	1	
	1977	LBN	1	1	
	1982	LBN	1	1	
	1984	LBN	0	1	
	1989	LBN	1	1	1
Liberia	All	LBR	Influential country*		
Myanmar	1970	MMR	1		0
Mauritania	1964	MRT	1	1	0
Nicaragua	1979	NIC	0	1	0
Oman	1967	OMN			1
	1968	OMN			1
Rwanda	1994	RWA	1	1	1
	1995	RWA	0	1	0
Swaziland	1974	SWZ	0	1	0
Zimbabwe	2009	ZWE	1	0	0

Notes: Outliers are dropped when the abs(log-per capita growth)>40% in a given year. Blank indicates missing data for that dataset. Liberia is dropped due to multiple influential observations due to the first and second Liberian Civil wars.

Appendix Table 2: Monte Carlo Estimates of Variance Components (True: sd(leader)=1.5%, sd(iid)=5%, homogenous)

	Simple Model (no country dummies)				**Full Model (with country dummies)**					
	SA-Method		**RE-Method**		**SA-Method**			**RE-Method**		
	SD(leader)	**sd(iid)**	**SD(leader)**	**sd(iid)**	**SD(leader)**	**sd(iid)**	**sd(CE)**	**SD(leader)**	**sd(iid)**	**sd(CE)**
No Country Effects	1.50%	5.00%	1.44%	5.00%	1.78%	5.00%	0.99%	1.52%	5.00%	0.99%
True sd(CE)=0	[0.11%]	[0.1%]	[0.43%]	[0.1%]	[0.14%]	[0.1%]	[0.07%]	[0.42%]	[0.1%]	[0.07%]
Country Effect Pvalue:*							76.85%			52.22%
With Country Effects	2.12%	5.00%	2.08%	5.00%	1.77%	5.00%	1.79%	1.49%	5.00%	1.80%
True sd(CE)=1.5%	[0.12%]	[0.09%]	[0.3%]	[0.09%]	[0.14%]	[0.09%]	[0.11%]	[0.45%]	[0.09%]	[0.11%]
Country Effect Pvalue:*							0.00%			0.00%

* P-value of test country effects (CE)=0. Note: Table presents monte carlo estimates of leader effects, where the real country X leader structure is used, but leader effects are drawn from a normal distribution with true SD 1.5%. Relative to Table 2 in the main text, in this table all of the iid errors are drawn from the same distribution (rather than having different distributions for autocrats and democrats). A successful method uncovers the "true" parameter of the leader effect of 1.5%. In the left panel the method has no country dummies, whereas on the right panel is the method with country dummies to detect country effects. iid errors are drawn from a mixture of normals, to replicate excess kurtosis in actual growth data. Standard deviations of bootstrap sample (across replications) are reported in brackets.

Appendix Table 3: Monte Carlo Estimates of Leader Effects (500 reps -- Autocrats & Democrats)

	PSI Mean*	PSI (SD)*	GAMMA Mean*	GAMMA (SD)*	Unbiased^	RMSE
	Panel 1: Autocracies					
	PSI Mean*	**PSI (SD)***	**GAMMA Mean***	**GAMMA (SD)***	**Unbiased^**	**RMSE**
	Panel 1A: Using Actual Variance Components					
No CE	0.26	0.17	0.00	0.00	1.00	1.29%
					[6.04%]	[0.03%]
With CE	0.23	0.13	0.69	0.09	0.99	1.32%
					[6.12%]	[0.03%]
	Panel 1B: Using SA Variance Components					
No CE	0.31	0.17	0.51	0.10	0.81	1.32%
	[3.9%]	[1.08%]	[3.86%]	[0.37%]	[12.75%]	[0.04%]
With CE	0.30	0.15	0.74	0.09	0.77	1.35%
	[3.96%]	[1.13%]	[3.01%]	[0.68%]	[13.16%]	[0.04%]
	Panel 1C: Using RE Variance Components					
No CE	0.24	0.13	0.55	0.09	1.76	1.35%
	[11.32%]	[5.07%]	[5.95%]	[1%]	[601.56%]	[0.07%]
With CE	0.23	0.12	0.76	0.08	1.39	1.37%
	[11.44%]	[4.95%]	[4.52%]	[1.51%]	[245.29%]	[0.07%]
	Panel 2: Democracies					
	PSI Mean*	**PSI (SD)***	**GAMMA Mean***	**GAMMA (SD)***	**Unbiased^**	**RMSE**
	Panel 2A: Using Actual Variance Components					
No CE	0.43	0.19	0.00	0.00	1.00	1.14%
					[6.88%]	[0.05%]
With CE	0.40	0.17	0.86	0.04	0.99	1.17%
					[6.59%]	[0.05%]
	Panel 2B: Using SA Variance Components					
No CE	0.43	0.18	0.53	0.08	0.98	1.15%
	[4.68%]	[0.59%]	[7.69%]	[0.67%]	[14.28%]	[0.05%]
With CE	0.42	0.17	0.87	0.04	0.96	1.17%
	[4.47%]	[0.53%]	[3.55%]	[0.91%]	[13.79%]	[0.05%]
	Panel 2C: Using RE Variance Components					
No CE	0.40	0.17	0.54	0.07	1.09	1.17%
	[9.3%]	[1.92%]	[8.35%]	[1.29%]	[34.11%]	[0.06%]
With CE	0.39	0.16	0.87	0.04	1.11	1.19%
	[9.03%]	[1.93%]	[4%]	[1.24%]	[101.67%]	[0.06%]

* Mean and SD across leaders (not MC draws of growth). ^ Unbiased =1 See notes for Table 3 (main text)

Appendix Table 4: Estimates of Variances Components - RE method (standard random effects)												
Method: RE	**Panel A: All leaders**				**Panel B: Autocrats**				**Panel C: Democrats**			
Dataset:	**PWT 7.1 Data**		**PWT9 Data**		**PWT 7.1 Data**		**PWT9 Data**		**PWT 7.1 Data**		**PWT9 Data**	
Leader SD	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.38%	0.00%	1.02%	0.72%	1.30%	0.81%
	[0.08%]	[0%]	[0.42%]	[0.16%]	[0.18%]	[0.03%]	[0.55%]	[0.18%]	[0.6%]	[0.43%]	[0.33%]	[0.35%]
iid Error	5.75%	5.50%	4.96%	4.73%	6.21%	5.93%	5.39%	5.13%	3.52%	2.67%	2.89%	2.22%
	[0.25%]	[0.24%]	[0.22%]	[0.22%]	[0.26%]	[0.26%]	[0.24%]	[0.23%]	[0.39%]	[0.32%]	[0.25%]	[0.25%]
Country SD	1.47%	1.26%	1.51%	1.35%	1.58%	1.36%	1.61%	1.46%	0.68%	0.39%	0.65%	0.40%
	[0.1%]	[0.11%]	[0.11%]	[0.14%]	[0.13%]	[0.23%]	[0.21%]	[0.25%]	[0.11%]	[0.06%]	[0.12%]	[0.07%]
(p-value country SD=0)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.69	0.14	0.33
YearXContinent FE	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes
Commodity Pr	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes	-	Yes
Obs / Leaders	6706 /1102		7147 / 1155		5214 / 783		5531 / 811		1492 / 319		1616 / 344	

Notes: This table presents estimates of the standard deviation of different variance componements of the model in Equation (1) and (2). The estimates are formed in two stages. First, one regresses the per capita real growth rate (by different datasets PWT7.1 or PWT9) on obervables: time by continent FE and/or country-specific commodity price indices (based on that country's exports over 2003-07). Then one collects the residuals and runs an regression of the residuals on country dummies, and decompose the error into a leader effect and a iid error, using a standard random effects estimator (re in stata). Democracies are countries with an average polity score above 7.5 and autocracies are all other countries. Standard errors are calculated using a country-level block bootstrap with 500 replications. Outliers with greater than 40% growth in a particular year (in absolute value) and Libera dropped.

Appendix Table 4C: Estimates of Variances Components -- World Development Indicators Data (SA and RE)												
	Panel A: All Leaders				**Panel B: Autocrats**				**Panel C: Democrats**			
	Method 1: SA (Random effects with unbalanced panel adjustment)											
Leader SD	1.74%	1.42%	1.42%	1.17%	1.89%	1.44%	1.44%	1.29%	1.63%	1.20%	1.16%	0.82%
iid Error	4.67%	4.45%	4.44%	4.14%	5.07%	4.83%	4.82%	4.48%	2.46%	1.86%	1.86%	1.41%
Country SD	1.58%	1.44%	1.46%	1.23%	1.73%	1.57%	1.59%	1.30%	0.66%	0.44%	0.43%	0.39%
	Method 2: SA (Standard Random effects)											
Leader SD	1.13%	0.32%	0.31%	0.00%	1.33%	0.00%	0.00%	0.00%	1.63%	1.11%	1.09%	0.61%
iid Error	4.67%	4.45%	4.44%	4.14%	5.07%	4.83%	4.82%	4.48%	2.46%	1.86%	1.86%	1.41%
Country SD	1.57%	1.44%	1.46%	1.21%	1.72%	1.57%	1.58%	1.29%	0.66%	0.44%	0.43%	0.39%
YearXContinent FE	-	Yes	Yes		-	Yes	Yes		-	Yes	Yes	
Commodity Pr	-	-	Yes		-	-	Yes		-	-	Yes	
TimeXSubregion				Yes				Yes				Yes
Obs / Leaders	Obs=6209; Leaders=1008				Obs=4900; Leaders=736				Obs=1309; Leaders=272			

Notes: This table presents the estimates of the standard deviation of different variance componements of the model in Equation (1) and (2). The estimates are formed in two stages. First, one regresses the per capita real growth rate (by different datasets PWT7.1, WDI, or PWT9) on obervables: time by continent FE and/or country-specific commodity price indices (based on that country's exports over 2003-07). Then we collect the residuals and run a regression of the residuals on country dummies, and decompose the error into a leader effect and a iid error, using the standard random effects estimator (re- without unbalanced panel adjustment) or a version with an unbalanced panel adjustment (sa). Democracies are countries with an average polity score above 7.5 and autocracies are all other countries. Outliers with greater than 40% growth in a particular year (in absolute value) and Libera dropped.

Appendix Table 4A: Estimates of Variances Components SA - Alternative Specifications

				Panel A: All leaders				
Method: SA		**PWT 7.1 Data**				**PWT9 Data**		
Leader SD	0.88%	1.09%	0.74%	0.84%	1.42%	1.58%	1.25%	1.38%
iid Error	5.52%	5.65%	5.16%	5.50%	4.74%	4.86%	4.40%	4.73%
Country SD	1.27%	1.44%	1.07%	1.25%	1.37%	1.51%	1.13%	1.37%
TimeXContinent FE	Yes	Year FE		Yes	Yes	Year FE		Yes
TimeXSubregion	-	Only	Yes		-	Only	Yes	
BB Comm. Pr & Civil war		-	-	Yes		-	-	Yes
Obs / Leaders		Obs=6706; Leaders=1102				Obs=7147; Leaders=1155		
				Panel B: Autocrats				
Method: SA		**PWT 7.1 Data**				**PWT9 Data**		
Leader SD	1.01%	1.35%	0.98%	0.97%	1.52%	1.80%	1.44%	1.48%
iid Error	5.95%	6.10%	5.55%	5.93%	5.14%	5.27%	4.77%	5.13%
Country SD	1.38%	1.54%	1.13%	1.36%	1.49%	1.63%	1.19%	1.49%
TimeXContinent FE	Yes	Year FE		Yes	Yes	Year FE		Yes
TimeXSubregion	-	Only	Yes		-	Only	Yes	
BB Comm. Pr & Civil war		-	-	Yes		-	-	Yes
Obs / Leaders		Obs=5214; Leaders=783				Obs=5531; Leaders=811		
				Panel C: Democrats				
Method: SA		**PWT 7.1 Data**				**PWT9 Data**		
Leader SD	1.01%	0.95%	0.68%	0.91%	1.03%	1.04%	0.61%	0.97%
iid Error	2.69%	3.22%	2.15%	2.64%	2.23%	2.62%	1.74%	2.21%
Country SD	0.41%	0.68%	0.36%	0.41%	0.42%	0.65%	0.34%	0.42%
TimeXContinent FE	Yes	Year FE		Yes	Yes	Year FE		Yes
TimeXSubregion	-	Only	Yes		-	Only	Yes	
BB Comm. Pr & Civil war		-	-	Yes		-	-	Yes
Obs / Leaders		Obs=1492; Leaders=319				Obs = 1616; Leaders=344		

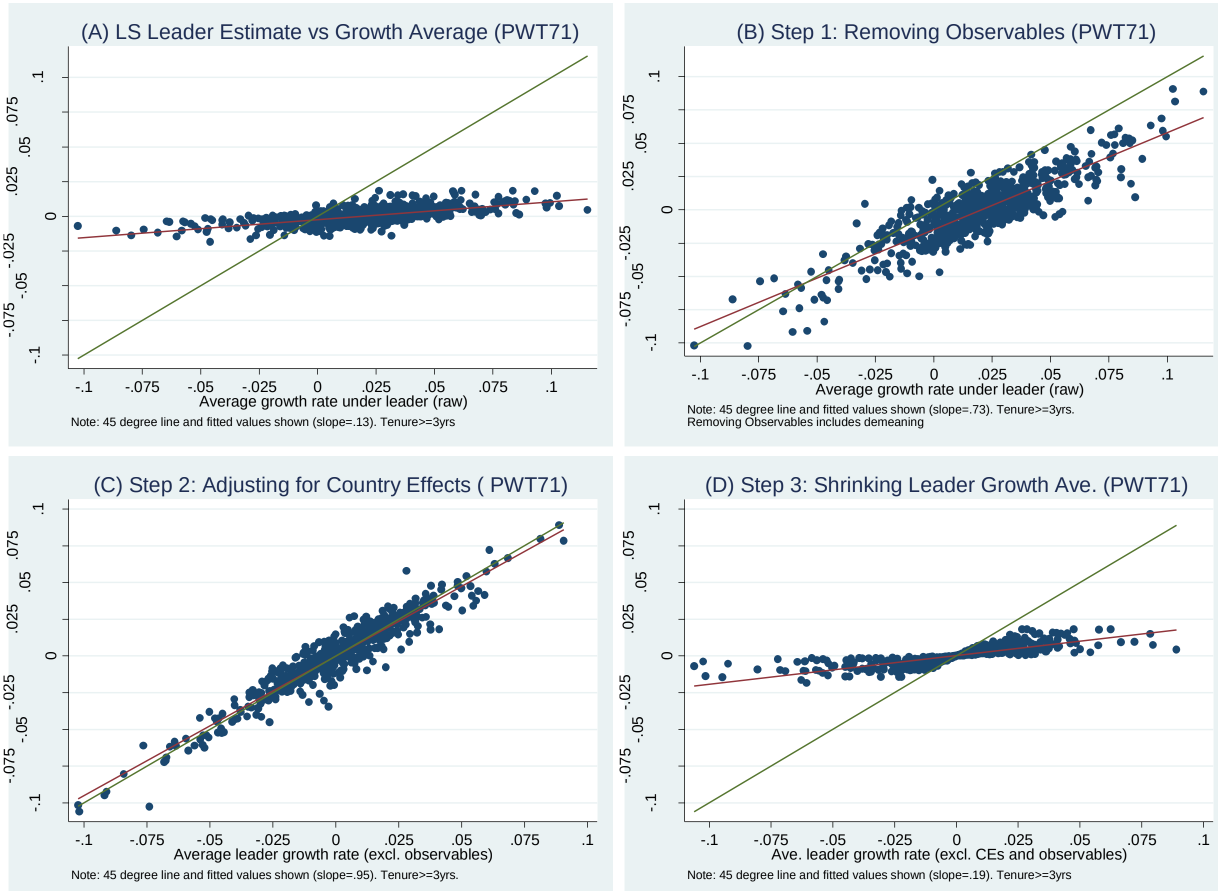
Notes: This table presents the estimates of the standard deviation of different variance componements of the model in Equation (1) and (2). The estimates are formed in two stages. First, one regresses the per capita real growth rate (by different datasets PWT7.1, WDI, or PWT9) on obervables: time by continent FE and/or country-specific commodity price indices (based on that country's exports over 2003-07). Then we collect the residuals and run an regression of the residuals on country dummies, and decompose the error into a leader effect and a iid error, using the unbalanced panel adjustmented random effects estimator (sa). Democracies are countries with an average polity score above 7.5 and autocracies are all other countries. Outliers with greater than 40% growth in a particular year (in absolute value) and Libera dropped.

Appendix Table 4B: Estimates of Variances Components RE - Alternative Specifications

	Panel A: All leaders							
Method: RE	**PWT 7.1 Data**				**PWT9 Data**			
Leader SD	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
iid Error	5.52%	5.65%	5.16%	5.50%	4.74%	4.86%	4.40%	4.73%
Country SD	1.27%	1.45%	1.08%	1.26%	1.34%	1.49%	1.09%	1.36%
TimeXContinent FE	Yes	Year FE		Yes	Yes	Year FE		Yes
TimeXSubregion	-	Only	Yes		-	Only	Yes	
BB Comm. Pr & Civil war		-	-	Yes		-	-	Yes
Obs / Leaders	Obs=6706; Leaders=1102				Obs=7147; Leaders=1155			
	Panel B: Autocrats							
Method: RE	**PWT 7.1 Data**				**PWT9 Data**			
Leader SD	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
iid Error	5.95%	6.10%	5.55%	5.93%	5.14%	5.27%	4.77%	5.13%
Country SD	1.37%	1.56%	1.14%	1.36%	1.45%	1.61%	1.15%	1.47%
TimeXContinent FE	Yes	Year FE		Yes	Yes	Year FE		Yes
TimeXSubregion	-	Only	Yes		-	Only	Yes	
BB Comm. Pr & Civil war		-	-	Yes		-	-	Yes
Obs / Leaders	Obs=5214; Leaders=783				Obs=5531; Leaders=811			
	Panel C: Democrats							
Method: RE	**PWT 7.1 Data**				**PWT9 Data**			
Leader SD	0.75%	0.00%	0.00%	0.74%	0.84%	0.61%	0.00%	0.81%
iid Error	2.69%	3.22%	2.15%	2.64%	2.23%	2.62%	1.74%	2.21%
Country SD	0.42%	0.70%	0.41%	0.41%	0.42%	0.67%	0.39%	0.42%
TimeXContinent FE	Yes	Year FE		Yes	Yes	Year FE		Yes
TimeXSubregion	-	Only	Yes		-	Only	Yes	
BB Comm. Pr & Civil war		-	-	Yes		-	-	Yes
Obs / Leaders	Obs=1492; Leaders=319				Obs = 1616; Leaders=344			

Notes: This table presents the estimates of the standard deviation of different variance componenements of the model in Equation (1) and (2). The estimates are formed in two stages. First, one regresses the per capita real growth rate (by different datasets PWT7.1, WDI, or PWT9) on obervables: time by continent FE and/or country-specific commodity price indices (based on that country's exports over 2003-07). Then we collect the residuals and run an regression of the residuals on country dummies, and decompose the error into a leader effect and a iid error, using the standard random effects estimator (re) - without unbalanced panel adjustment. Democracies are countries with an average polity score above 7.5 and autocracies are all other countries. Outliers with greater than 40% growth in a particular year (in absolute value) and Libera dropped.

Appendix Figure 3: Least Squares Leader Effect Estimates – Pooled Sample (PWT7.1)



Appendix Table 6: Best 25 leaders according to PWT7.1 least squares estimates (tenure 4yrs+, country-specific iid SD)

		Leader Rank (PWT71)		Leader Rank (PWT9)		LS Leader Effect		Growth Average		Demo-	
Leader name	Country	Least Sq	Growth Ave	Least Sq	Growth Ave	PWT 7.1	PWT9	PWT 7.1	PWT9	cracy	Tenure
Raab	AUT	1		21		1.8%	1.8%	6.2%	6.1%	1	8
Reagan	USA	2		24		1.8%	1.7%	2.6%	2.6%	1	8
Medici	BRA	3	8	4	8	1.8%	2.6%	8.4%	8.1%	0	5
Hun Sen	KHM	4		2		1.8%	2.8%	5.8%	5.2%	0	25
Soares	PRT	5		11		1.6%	2.3%	3.7%	3.3%	0	10
Kubitschek	BRA	6		5		1.6%	2.6%	5.9%	5.9%	0	5
Manmohan Singh	IND	7	21	10		1.6%	2.3%	6.8%	6.0%	1	10
Museveni	UGA	8		7		1.5%	2.5%	3.1%	3.4%	0	28
Chissano	MOZ	9		16		1.5%	2.0%	4.1%	4.3%	0	19
Cristiani	SLV	10		13		1.5%	2.2%	3.4%	4.1%	0	5
Khama	BWA	11	1	3	1	1.5%	2.8%	10.2%	9.6%	0	15
Mubarak	EGY	12				1.5%	1.2%	4.2%	3.4%	0	30
Zhivkov	BGR	13				1.4%	1.7%	5.0%	5.1%	0	19
Chun Doo Hwan	KOR	14	15	1	6	1.4%	3.3%	7.7%	8.3%	0	8
Gonzalez Marquez	ESP	15				1.3%	1.6%	2.5%	2.4%	0	14
Santer	LUX	16				1.3%	1.3%	4.7%	4.3%	1	11
Vargas	COL	17				1.2%	1.7%	1.6%	2.2%	0	4
Ikeda	JPN	18	3		4	1.2%	1.3%	8.9%	8.6%	1	4
Razak	MYS	19	10	6	15	1.2%	2.5%	8.1%	7.7%	0	6
Manning	TTO	20				1.2%	0.8%	6.1%	3.7%	1	13
Deng Xiaoping	CHN	21	9			1.1%	0.9%	8.2%	5.5%	0	17
Stroessner	PRY	22				1.1%	0.9%	2.3%	2.3%	0	35
Sato	JPN	23	12		14	1.1%	1.2%	8.0%	7.7%	1	8
Papadopoulos	GRC	24	6		7	1.1%	1.6%	8.5%	8.2%	0	6
Zia	PAK	25		12		1.1%	2.3%	3.3%	3.3%	0	11

Notes: Blank indicates rank outside the top 25. Tenure to 2014. See notes in main text Table 6

Appendix Table 7: Worst 25 leaders according to PWT7.1 least squares estimates (tenure 4yrs+; het iid SD)- Rank from Last

		Leader Rank (PWT71)		Leader Rank (PWT9)		LS Leader Effect		Growth Average		
Leader name	Country	Least Sq	Growth Ave	Least Sq	Growth Ave	PWT 7.1	PWT9	PWT 7.1	PWT9	Tenure
Amin	UGA	1	16	1	10	-1.8%	-3.8%	-4.6%	-5.0%	8
Plaek Pibulsongkram	THA	2	24	5	24	-1.6%	-2.4%	-2.9%	-2.9%	7
Ochirbat	MNG	3	7	3		-1.5%	-3.0%	-6.0%	-2.1%	7
Prodi	ITA	4				-1.4%	-1.0%	0.5%	0.8%	4
Chen Shui-bian	TWN	5		23		-1.4%	-1.5%	3.2%	3.6%	8
G.W. Bush	USA	6				-1.4%	-1.0%	0.3%	0.7%	8
Machel	MOZ	7		12	25	-1.4%	-2.0%	-2.6%	-2.6%	12
Laurent Gbagbo	CIV	8		15		-1.4%	-1.7%	-1.4%	-1.2%	11
Lon Nol	KHM	9	3	2	2	-1.4%	-3.0%	-8.0%	-9.1%	5
Yameogo	BFA	10		8		-1.3%	-2.2%	-2.0%	-2.3%	7
Carazo Odio	CRI	11	13		21	-1.1%	-1.4%	-4.8%	-3.7%	4
Patterson	JAM	12				-1.1%	-1.1%	-0.1%	0.3%	14
Mao Tse-Tung	CHN	13		22		-1.1%	-1.5%	2.7%	1.6%	24
Fitzgerald	IRL	14				-1.1%	-0.8%	0.1%	1.2%	6
Rodriguez Zapatero	ESP	15				-1.1%	-1.3%	-0.3%	0.0%	7
Chambers	TTO	16	2		7	-1.0%	-1.4%	-8.6%	-5.5%	5
Alia	ALB	17	8	7	4	-1.0%	-2.2%	-5.8%	-6.7%	7
Sukarno	IDN	18		16		-1.0%	-1.7%	0.0%	0.0%	6
Galvez	HND	19	14	6	9	-1.0%	-2.3%	-4.7%	-5.0%	4
Senghor	SEN	20				-1.0%	-1.4%	-1.0%	-1.1%	20
Nehru	IND	21				-1.0%	-1.3%	2.3%	2.2%	14
MacMillan	GBR	22				-1.0%	-0.9%	1.8%	1.8%	6
Pizano	COL	23				-1.0%	-1.0%	0.6%	1.2%	4
van Agt	NLD	24				-1.0%	-1.1%	-0.5%	0.2%	5
Ayatollah Khomeini	IRN	25	12	19	8	-1.0%	-1.6%	-5.1%	-5.4%	10

Notes: Blank indicates rank outside the bottom 25. "Rank" is from last of 609 leaders with tenure (and PWT7.1 data) of at least 4 ye

Appendix Table 8: Best 25 leaders according to WDI least squares estimates (tenure 4yrs+, country-specific iid SD)

		Leader Rank (WDI)		Leader Rank (PWT9)		LS Leader Effect		Growth Average		Demo-	
Leader name	**Country**	**Least Sq**	**Growth Ave**	**Least Sq**	**Growth Ave**	**WDI**	**PWT9**	**Wdi**	**PWT9**	**cracy**	**Tenure**
Veiga	CPV	1	3			3.8%	1.4%	9.3%	4.6%	0	9
Chun Doo Hwan	KOR	2	5	1	6	3.4%	3.3%	8.4%	8.3%	0	8
Khama	BWA	3	2	3	1	3.0%	2.8%	9.4%	9.6%	0	15
Rodriguez Lara	ECU	4	11	20	11	3.0%	1.8%	7.5%	8.0%	0	4
Manmohan Singh	IND	5		10		2.8%	2.3%	5.9%	6.0%	1	10
Medici	BRA	6	9	4	8	2.8%	2.6%	8.0%	8.1%	0	5
Lee Kuan Yew	SGP	7		8		2.1%	2.4%	6.3%	6.3%	0	30
Chissano	MOZ	8		16		2.1%	2.0%	5.0%	4.3%	0	19
Cristiani	SLV	9		13		2.1%	2.2%	4.4%	4.1%	0	5
Zia	PAK	10		12		2.1%	2.3%	3.4%	3.3%	0	11
Lopez Portillo	MEX	11		25		2.1%	1.7%	3.8%	3.5%	0	6
Reagan	USA	12		24		1.9%	1.7%	2.6%	2.6%	1	8
Vargas	COL	13				1.9%	1.7%	2.6%	2.2%	0	4
Soares	PRT	14		11		1.8%	2.3%	3.5%	3.3%	0	10
Verwoerd	ZAF	15				1.8%	0.6%	3.7%	2.7%	0	8
A. Nastase	ROU	16	16		23	1.8%	0.9%	7.1%	6.6%	0	4
Elbegdorj	MNG	17	4	14	3	1.8%	2.2%	8.7%	8.7%	0	5
Stroessner	PRY	18				1.8%	0.9%	3.3%	2.3%	0	35
Hasina Wazed	BGD	19		18		1.7%	1.9%	3.8%	4.0%	0	10
Prem	THA	20				1.7%	1.1%	4.8%	4.8%	0	8
Panday	TTO	21	23		25	1.7%	1.2%	6.6%	6.5%	1	6
Museveni	UGA	22		7		1.7%	2.5%	3.2%	3.4%	0	28
Sato	JPN	23	24		14	1.7%	1.2%	6.6%	7.7%	1	8
Traian Basescu	ROU	24				1.7%	0.9%	3.4%	3.6%	0	10
Hun Sen	KHM	25		2		1.7%	2.8%	5.8%	5.2%	0	25

Notes: Blank indicates rank outside the top 25. Tenure to 2014. See notes in main text Table 6

Appendix Table 9: Worst 25 leaders according to WDI least squares estimates (tenure 4yrs+; het iid SD)- Rank from Last

		Leader Rank (WDI)		Leader Rank (PWT9)		LS Leader Effect		Growth Average		Demo-	
Leader name	**Country**	**Least Sq**	**Growth Ave**	**Least Sq**	**Growth Ave**	**Wdi**	**PWT9**	**Wdi**	**PWT9**	**cracy**	**Tenure**
Dimitris Christofias	CYP	1	12		19	-3.2%	-0.7%	-3.7%	-3.9%	1	5
Gouled Aptidon	DJI	2	9	10	16	-2.6%	-2.0%	-4.1%	-4.1%	0	23
Ochirbat	MNG	3		3		-2.5%	-3.0%	-2.1%	-2.1%	0	7
Alia	ALB	4	-1	7	4	-2.3%	-2.2%	-7.0%	-6.7%	0	7
Machel	MOZ	5	3	12	25	-2.2%	-2.0%	-6.2%	-2.6%	0	12
Fahd	SAU	6	16			-1.9%	-1.1%	-2.6%	-2.3%	0	14
Chambers	TTO	7	5		7	-1.9%	-1.4%	-5.5%	-5.5%	1	5
Sukarno	IDN	8		16		-1.9%	-1.7%	-0.6%	0.0%	0	6
Alvarez Armalino	URY	9	4		13	-1.8%	-1.0%	-5.8%	-4.5%	0	4
Laurent Gbagbo	CIV	10		15		-1.8%	-1.7%	-1.4%	-1.2%	0	11
Bandaranaike, S	LKA	11				-1.8%	-0.8%	1.6%	1.8%	0	12
Mao Tse-Tung	CHN	12		22		-1.7%	-1.5%	2.0%	1.6%	0	24
Mobutu	COD	13	14	17	22	-1.7%	-1.7%	-3.4%	-3.4%	0	32
Carazo Odio	CRI	14	11		21	-1.6%	-1.4%	-3.7%	-3.7%	1	4
Acheampong	GHA	15				-1.6%	-1.2%	-1.7%	-1.3%	0	6
Anastasio Somoza Del	NIC	16	22	9		-1.6%	-2.0%	-2.4%	-2.5%	0	12
Manley	JAM	17	17			-1.6%	-1.0%	-2.6%	-1.4%	1	11
Khalifa Al Nahayan	ARE	18	7		12	-1.6%	-1.5%	-4.7%	-4.6%	0	10
Lee Myung Bak	KOR	19				-1.5%	-1.0%	2.6%	2.6%	0	5
Roh Moo Hyun	KOR	20				-1.5%	-1.3%	3.9%	3.7%	0	5
Ferenc Gyurcsany	HUN	21		18		-1.4%	-1.6%	0.7%	0.7%	0	5
Gandhi, I.	IND	22				-1.4%	-0.9%	2.1%	2.3%	1	15
Kaunda	ZMB	23				-1.4%	-1.0%	-1.2%	-0.7%	0	28
Sampaio	PRT	24				-1.4%	-1.2%	1.8%	1.9%	0	10
Calderon	MEX	25				-1.4%	-1.5%	0.6%	0.6%	0	6

Notes: Blank indicates rank outside the bottom 25. "Rank" is from last of 549 leaders with tenure (and WDI data) of at least 4 years.

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). As in Jones and Olken (2005), in the case there are multiple leaders in a year we keep the leader who ended his/her tenure in that year and started their tenure earliest – ie the leader who is in office on 1 January "gets" that year.[46]

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 a democracies having an average polity score >=7.5, and autocracies <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 chain linked 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 and Liberia (see Section 2 and Appendix table 1a). We drop countries with less than 30 years of growth data (combined across all our data sources).[47]

---

[46] 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.

[47] This also includes data from PWT8.1, even though for brevity we do not include those results in the paper.