St Comp Int Dev (2017) 52:261–277
DOI 10.1007/s12116-015-9192-4

The Myth of the Democratic Advantage

Rory Truex[1]

Published online: 4 August 2015

Abstract Existing research points to a democratic advantage in public good provision. Compared to their authoritarian counterparts, democratically elected leaders face more political competition and must please a larger portion of the population to stay in office. This paper provides an impartial reevaluation of the empirical record using the techniques of global sensitivity analysis. Democracy proves to have no systematic association with a range of health and education outcomes, despite an abundance of published empirical and theoretical findings to the contrary.

Keywords Democracy · Authoritarianism · Regime type · Public good provision · Sensitivity analysis

Democracies are thought to have a comparative advantage in public good provision. The theoretical reasoning is straightforward. To stay in office, elected leaders must please a large portion of the population, and providing schools, water, electricity, access to vaccinations, and other basic necessities is usually the best way to garner that support. In contrast, authoritarian rulers generally have a less "encompassing interest" and must placate smaller "selectorates" and "winning coalitions." Rather than provide expensive public goods, they may prefer to buy support by distributing private rents (Bueno de Mesquita et al. 2003; Deacon 2009; McGuire and Olson 1996). Authoritarianism reduces political competition and access to information, which should also serve to worsen governance (Brown 1999; Lake and Baum 2001).

Replication data available at www.rorytruex.com.

Electronic supplementary material The online version of this article (doi:10.1007/s12116-015-9192-4) contains supplementary material, which is available to authorized users.

✉ Rory Truex
rtruex@princeton.edu

[1] Princeton University, Princeton, NJ, USA

The empirical record to date largely supports this logic (Deacon and Saha 2006), albeit with some exceptions (Mulligan et al. 2004; Ross 2006).[1] The standard approach is to investigate the conditional association between democracy and various public good indicators, and then to infer (either implicitly or explicitly) a causal relationship from the correlation. In their influential study on "the invisible hand of democracy," Lake and Baum (2001) examine the relationship between regime type and public health and education indicators. Democracy, as measured by Polity scores, proves associated with large and significant improvements in adult literacy, the primary pupil teacher ratio, enrollment rates, healthcare access, clean water access, the number of available physicians, the percentage of births attended by health personnel, immunization rates, infant mortality rates, crude death rates, and life expectancy. An earlier study by Brown (1999) finds that democracy increases primary school enrollment, although the effect may be conditional on income levels. Several studies show associations between democracy and lower infant mortality rates, longer life expectancies, and other health outcomes (Besley and Kudamatsu 2006; Boone 1996; Franco et al. 2004; Kudamatsu 2012; Moon and Dixon 1985; Przeworski et al. 2000; Wigley and Akkoyunlu-Wigley 2011; Zweifel and Navia 2000). Bueno de Mesquita et al. (2003) take the positive effect of democracy as given and focus on separating out the independent effects of the size of the winning coalition W, which is largest under democracy. They find a positive relationship between their measure of W and a broad range of public good indicators. A more recent inquiry by Deacon produces the familiar result: "dictatorial governments are found to provide public schooling, roads, safe water, public sanitation, and pollution control at levels far below democracies" (Deacon 2009: 241).

Closer inspection reveals that the empirical record on the "democratic advantage" is indeed quite muddled. Different studies employ different statistical models, include different covariate sets and lag structures, examine different years, and assess different public good provision measures. At best, the existing empirical record is incomplete. At worst, the entire literature could be distorted by publication bias, especially if researchers are presenting and submitting findings based on statistical significance, or if journals' publication decisions are similarly related to effect size (Gerber et al. 2001; Gerber and Malhotra 2008).

Establishing a causal inference in this setting requires that (a) an association between regime type and public good provision exists and (b) there is enough evidence, both theoretical and empirical, to suggest is a causal arrow between the two. We know the literature tends to fall short in terms of causal identification. The core contribution of this paper is to investigate the associational evidence, which is the easier of the two conditions. If the association proves weak, this would raise serious doubts about any causal relationship.

A number of econometricians have pushed for the use of more comprehensive robustness checks. In his classic piece, "Sensitivity Analysis Would Help," Leamer (1985) argues that the advancement of knowledge requires a higher standard for

[1] Ross (2006) shows that once missing data problems are accounted for, there is no significant relationship between democracy and infant mortality rates. The expenditures literature has two reported findings that contradict the democratic advantage argument. Mulligan et al. (2004) consider the relationship between regime type and public policy broadly defined, and find that there are not major differences between autocracies and democracies in terms of education spending.

empirical research. Empirical studies must show that minor changes in the list of assumptions—different covariates, lag structures, measurement of key variables, etc.—do not alter the core inferences. If a very specific set of assumptions is required to produce a finding, the inferences are simply "too fragile to be believed."

Leamer (1985) goes on to propose a "global sensitivity analysis" approach, which involves systematically examining a broad range of plausible assumptions to assess "inferential sturdiness." Levine and Renelt (1992) and Sala-I-Martin (1997) extend this reasoning in their studies of the determinants of economic growth, the latter running upwards of two million models to place different independent variables on a spectrum of robustness. Recent studies by political scientists have begun to subject well-established empirical relationships to similar scrutiny. Hegre and Sambanis (2006) adopt Sala-I-Martin's approach to assess the growing literature on civil war onset, and Gassebner et al. (2013) systematically reevaluate the determinants of democratic transitions and stability. Others have conducted comprehensive robustness checks of central topics like the relationship between economic shocks and conflict and the democratic peace (Bazzi and Blattman 2011; Dafoe 2011). The growth of these types of inquiries suggests an increasing appreciation of Leamer's central message: "a fragile inference is not worth taking seriously" (Leamer 1985: 308).

This paper will not develop a new theory of authoritarian politics, nor create a new conceptual distinction, nor build a new original dataset of public good measures. Its sole purpose is to test the sturdiness of the cross-sectional association between regime type and public good provision. Not all relationships merit this sort of attention, but within comparative politics, the association between democracy and health and education outcomes stands out as one of the most prominent inferences in the subfield.

Following the prescriptions of these earlier studies, I examine a range of possible plausible assumptions—different years, different possible lag structures, different conceptions of democracy, different measures of public good provision, different sample restrictions/approaches to missingness, and most importantly, different covariate sets (Leamer 1985). Around 300,000 cross-section models were estimated in total.

The sensitivity analysis offers some evidence that democracies may be better at promoting access to clean water and some very weak evidence that they foster lower infant mortality rates and lengthen life expectancies. There is effectively no evidence, however, that they are better at fostering primary school enrollment, fostering secondary school enrollment, reducing pupil-teacher ratios in primary schools, reducing pupil-teacher ratios secondary schools, promoting the completion of primary school, providing measles immunization, providing DPT immunization, promoting access to physicians, promoting access to hospital beds, promoting access to sanitation, and reducing crude death rates. Only one out of the fourteen public good outcomes passes standards for robustness under the global sensitivity analysis framework.

It is important to note that these findings are limited to public good outcomes: school enrollment rates, vaccinations received, clean water access, and so forth. A related body of research points to a positive democracy effect on public spending (Avelino et al. 2005; Brown and Hunter 1999, 2004; Kaufman and Segura-Ubiergo 2001; Stasavage 2005). In their review of the literature, Deacon and Saha (2006) warn against relying too heavily on expenditure measures, which may be only loose proxies for public good outcomes. Spending data may be inflated by corruption or pork barrel projects, distorted by cross-country differences in factor prices, or muddled by decentralization and other fiscal reforms (Kaufman and Segura-Ubiergo 2001). I share these reservations and will focus my analysis on outcome measures. There is little evidence that democracies perform better in terms of health and education outcomes, although it may still be possible that they spend more.

Global Sensitivity Analysis

Leamer's (1985) original method for more systematic robustness checks—"global sensitivity analysis" in his terminology—involves the estimation of the "extreme bounds" on a coefficient of interest.[2] Analysts should specify and estimate a wide range of plausible models, evaluating the range of inferences across different functional forms, covariate sets, measurement schemes, error structures, and so forth. Levine and Renelt (1992) develop and implement a variant of Leamer's methodology in their global sensitivity analysis of the determinants of economic growth. They specify equations of the following type:

$$Y = \beta_{\mathrm{i}} I + \beta_{\mathrm{m}} M + \beta_{\mathrm{z}} Z + u$$

Here, Y is the dependent variable (economic growth), I is a vector of variables that is generally considered to be part of the "true model," M is the variable of theoretical interest, and Z is a subset of variables selected from a larger pool of variables that might be part of the "true model." They estimate $\hat{\beta}_{\mathrm{m}}$ for all possible combinations of up to three variables in Z. With these estimates in hand, they then determine the extreme upper and lower bounds for the true value of β_{m}, with the upper (lower) bound defined as the maximum (minimum) estimate of β_{m} plus (minus) two standard errors. The effect of M is deemed robust only if both the extreme bounds are of the same sign. Note that this means that even if millions of regressions are estimated, a variable may fail to pass robustness if its coefficient fails to reach significance in just a single regression.

Unsurprisingly, Levin and Renelt's inquiry yielded very few robust results, leading others to develop methods with more reasonable standards. Sala-I-Martin's (1997) approach shares Levin and Renelt's method of choosing the range of plausible alternative models, but differs in its means of aggregating and evaluating the results. The analyst combines the set of plausible J estimates in such a way as to measure the proportion on either side of zero, or CDF(0). If 75 % of estimates for a given covariate exceed zero, it is more likely to be correlated with the dependent variable than a covariate with only 50 % of estimates exceeding zero (Sala-I-Martin 1997).[3] The primary advantage of this CDF(0) metric is that it can be used to put variables and correlations on a spectrum of robustness.[4]

[2] The discussion here follows closely that of Hegre and Sambanis (2006).

[3] Sala-I-Martin 1997, 179. Technically, the appropriate method for calculating the cumulative distribution function is still under debate. If the estimates appear normally distributed, Sala-I-Martin recommends taking a weighted mean of the estimates and their variances. The idea behind the weighting scheme is to give more weight to estimates from models with better fit, under the assumption that they are more likely to be the "true model." This weighting is not strictly necessary, and it has recently been criticized on the grounds that goodness-of-fit metrics are generally incomparable across models with different numbers of covariates and observations (Gassebner et al. 2013). For this reason, I will calculate the simple unweighted means for the estimates.

[4] The standard approach is to calculate the mean $\overline{\beta}_{\mathrm{m}}$ and variance $\overline{\sigma}^2_{\mathrm{m}}$, and then compute the cumulative distribution function using the standard normal approximation. Yet, given that my estimates are unweighted, the simpler method is to just examine the empirical distribution and avoid the normal approximation entirely.

Sala-I-Martin recommends that the benchmark of 0.95 be used as the threshold of "significance." That is, if the cumulative distribution function suggests that 95 % of the coefficient estimates are less than 0, then we would say that variable has a robust negative association with the dependent variable. We would use the symmetric logic for considering a positive association. For ease of exposition, I will refer to this as the CDF(0) metric throughout, but note that 1-CDF(0) was used for all positive association.

The CDF(0) metric focuses on the sign of the set of estimates, but we should also be interested in their magnitude. It is possible for an association to pass the CDF(0)>0.95 threshold, but for all the individual estimates to fail to reach significance, as in the case where all the estimates are weakly negative. This analysis presents two additional metrics of interest. The first is the fraction of estimates that reject the one-sided null hypothesis of no association at the 5 % significance level. There is no current benchmark for what would constitute a robust relationship under this metric, so I leave its interpretation open to the reader. In general, the greater fraction of estimates that are significant, the more we should believe there is a relationship in the data. Second, I present the average *p* value of the estimates, which is calculated using the *t* statistic from the ratio of the mean estimate over the mean standard error, $\overline{\beta}_{\mathrm{m}}/\overline{\sigma}_{\mathrm{m}}$. This measure is effectively a one-sided test of significance, and it gives a sense of whether the association is likely more than just statistical noise (Hegre and Sambanis 2006).

Applying Global Sensitivity Analysis

The core empirical findings on the relationship between regime type and public good provision are primarily from cross-section analyses, although some authors also estimate TSCS specifications for a few dependent variables. The cross-section models take a familiar form:

$$Y_{\mathrm{i}} = \alpha + \gamma D_{\mathrm{i}} + \beta X_{\mathrm{i}} + \varepsilon_{\mathrm{i}}$$

Here, *Y* is a measure of some public good, *D* is a measure of democracy, and *X* is a vector of country-level covariates. Robust standard errors account for heteroskedasticity in the error term. For some specifications, *D* may be lagged to account for the possibility that regime type may not have instantaneous effects on all public goods. Assuming large values of *Y* indicate strong public good provision, the "democratic advantage" hypothesis suggests that $\gamma>0$. Some of the dependent variables used in the analysis are actually "public bads" (infant mortality rates, pupil teacher ratios, etc.), so better governance implies that $\gamma<0$. Throughout the paper, these variables will be indicated with a (−) symbol to avoid confusion.

As Leamer (1985) recommends, we should seek to probe existing findings on as many dimensions as possible. Given problems of data coverage, this note will only present estimates from cross-section models, as has been the focus in existing research. The sensitivity analysis will simultaneously explore robustness across the dependent variable, the covariate set, the lag structure of the model, the measurement of democracy, and the year of analysis.

There is no single measure of "public good provision," but existing studies have employed a wide range of possible outcome measures in health and education. I focus my analysis on 14 frequently used dependent variables. For education, I examine the following: the pupil teacher ratio in primary (*prim.pupils*) and secondary schools (*sec.pupils*), the gross enrollment ratios in primary (prim.enroll) and secondary schools (*sec.enroll)*, and the primary completion rate (prim.completion). For health outcomes, I examine the following: overall access to water (water); overall access to adequate sanitation (sanitation); immunization rates for diphtheria, pertussis, and tetanus (dpt.im) and measles (measles.im); the number of hospital beds (beds) and physicians (physicians) per one thousand people; life expectancy (lifeexpectancy); the crude death rate (death); and the infant mortality rate (infmort). Together, these variables comprise the most commonly used health and education outcome measures in existing research. Table 1 shows the definitions and some descriptive statistics, including information on missingness (which is severe for some of the dependent variables).

For the covariate set, I follow the approach originally developed by Levine and Renelt (1992), choosing to keep a set of covariates *I* fixed and exploring robustness across the inclusion of covariates *Z* we are less sure about. In the public good provision literature, three concepts are consistently included as covariates across studies: logged income per capita (*logincome*), the total population (*population*), and the percentage of the population that is living in urban areas (urban).[5] If we had the power to randomly assign countries to regime type, at a minimum, we would want to observe balance across our treatment and control groups on these characteristics. They represent the "fixed" or "core" set of covariates *I*.

In terms of other covariates, a number have been suggested and employed. For my sensitivity analysis, I run all possible models that include up to four of the following additional concepts: (1) land area (area), (2) population density (popdens), (3) region indicators (africa, asia, ceurope, meast, nam, sam, and scan), (4) proportion of the population under age 14 and over 65 (pop14 and pop14), (5) percentage Muslim and Catholic (muslim and catholic), (6) aid as a percentage of national income (aid), and (7) ethnolinguistic fractionalization (elf). All of these covariates are plausible confounders for the relationship between regime type and public good provision, but they have garnered less agreement in the literature. Given that there are seven additional concepts, this "all possible combinations" approach yields a total of 98 models (7 with one additional concept, 21 with two additional concepts, 35 with three additional concepts, and 35 with four additional concepts).[6] The models are limited to just a few additional concepts to avoid problems of multicollinearity (Hegre and Sambanis 2006; Levine and Renelt 1992; Sala-I-Martin 1997). Table 2 below shows the balance of these covariates (excluding the region indicators) across democracies and non-democracies. We see that

[5] Lake and Baum (2001) actually include GNP per capita, not the logged GDP per capita, but graphical analysis shows that the logged model better approximates a linear relationship. Deacon (2009) also moves back and forth between various income and logged income measures. The replication analysis will employ logged GDP per capita.

[6] To calculate the total number of models, we can employ the combinations formula, $n!/(r!(n-r)!)$, where n is the total number of unique elements, and r is the number of unique elements in the combination. For this analysis, there are 21 possible models that have $2(r)$ out of the $7(n)$ concepts.

Table 1 Dependent variable descriptions

Variable	Description	Mean	SD	Missing (%)
prim.pupils (−)	Primary pupil-teacher ratio	27.64	16.82	31.9
sec.pupils (−)	Secondary pupil-teacher ratio	17.70	11.07	50.3
prim.enroll (+)	Primary gross enrollment ratio	103.91	13.99	15.7
sec.enroll (+)	Secondary gross enrollment ratio	77.66	28.21	24.9
prim.comp (+)	Primary completion rate (% of relevant age group)	86.24	18.87	44.4
dpt.im (+)	Immunization, DPT (% of children 12–23 months)	88.33	12.14	16.9
measles.im (+)	Immunization, measles (% of children 12–23 months)	87.15	13.24	18.0
water (+)	Improved water source (% of pop with access)	85.56	16.89	40.3
infmort (−)	Infant mortality rate (per 1000 live births)	31.44	30.12	1.6
death (−)	Crude death rate (per 1000 people)	9.10	3.63	0.4
sanitation (+)	Improved sanitation facilities (% of pop with access)	69.96	31.1	41.7
physicians (+)	Physicians (per 1000 people)	1.90	1.56	55.6
beds (+)	Hospital beds (per 1000 people)	3.82	2.68	63.2
lifeexpect (+)	Life expectancy at birth (years)	68.38	10.51	0.4

All variables drawn from the World Bank's World Development Indicators. The summary statistics reflect the data from 2009. The fraction missing is calculated over the entire analysis period

democracies are average are more developed and urbanized, and govern older populations, among other differences.

Like any cross-national regressions, these types of models are not on strong footing with respect to causal inference due to standard endogeneity critiques. It is possible that other factors could be confounding the relationship between regime type and public good provision, and no amount of conditioning will ever make us believe that the "democracy treatment" was as-if randomly assigned. Plausible instruments are also in short supply. As stated in the introduction, this paper pursues the more modest goal of investigating the strength of the conditional association between democracy and public good provision. If the association fails to hold, this would raise serious doubts about any causal relationship.

With respect to lag structure, it is unclear whether regime type has an instantaneous effect on public good provision or whether the effect may take a few years to materialize. Lake and Baum (2001) present estimates from the "optimal lag" that produces the largest effect for democracy, but this naturally biases the results in favor of the democratic advantage hypothesis. My analysis examines robustness across a 0- to 5-year lag structure. Whenever the democracy indicator D is lagged, the vector of covariates X will also be lagged to avoid biases induced by conditioning on post-treatment variables. The substantive findings do not change appreciably across different assumptions on the structure.

The analysis also considers the robustness of the finding across two different conceptions of democracy. Although I agree with the criticisms levied against the Polity score index (polity), it represents the primary measurement of democracy used in existing studies and will therefore be part of the sensitivity analysis (Coppedge et al. 2011;

Table 2 Covariates and balance statistics

Variable	Democracies		Non-democracies	
	Mean	SD	Mean	SD
logincome	8.67	1.61	7.64	1.32
population	5.34e+07	2.93e+16	4.51e+07	1.75e+08
urban	61.3	21.0	50.2	22.6
area	748,525	1,807,550	1,049,007	2,462,762
popdens	139.4	166.7	189.6	937.8
pop14	25.5	10.4	33.6	9.4
pop65	9.9	5.6	4.6	2.7
muslim	11.5	24.1	42.3	41.1
catholic	41.0	39.1	14.0	21.6
aid	5.4	9.1	6.4	7.7
elf	0.42	0.27	0.51	0.26

All covariates drawn from the World Bank's World Development Indicators. The summary statistics reflect the data from 2009. Democracy categorization uses GWF data

Munck and Verkuilen 2002).[7] In line with common practice, I subtract the Polity IV AUTOC score from the DEMOC score to produce a continuous scale from −10 to 10, with higher scores indicating more democratic governance.

I will also take advantage of new data on regime type gathered by Geddes, Wright, and Franz (GWF) (2012) to construct a simple binary measure of democracy/authoritarianism. The GWF data codes country-years into several distinct categories: autocratic, democratic, ruled by a provisional government, not independent, occupied by foreign troops, ruled by multiple governments, and no government at all. A regime spell is deemed autocratic if the executive achieved power through "undemocratic" means, which refers to anything besides "direct, reasonably fair, competitive elections in which at least ten percent of the total population was eligible to vote or an indirect election by a body, at least 60 % of which was elected in direct, reasonably fair, competitive elections; or constitutional succession to a democratically elected executive." The democracy coding follows the converse rule, and the dataset also separates out autocracy/democracy from periods of provisional government, foreign occupation, or warlordism. Country-years that were neither democratic nor autocratic are excluded from the sensitivity analysis. This results in the removal of about 125 country-years (~2.5 % of observations) that would have been included had standard binary datasets been used (Cheibub et al. 2010). The core results of the sensitivity analysis are not sensitive to this exclusion, as shown in the Online Appendix.

[7] This measure, which is a sum of five weighted ordinal scores on governance sub-measures (competitiveness of participation, regulation of participation, competitiveness of executive, openness of executive recruitment, constraints on executive), has been strongly criticized as including redundant components, omitting key aspects of democracy, and relying on a "convoluted aggregation rule" that has no theoretical justification. As a consumer of these indices, it is difficult to grasp what a unit change in a country's Polity score means in terms of reforms on the ground.

There is an additional concern that existing research is too selective in the year of analysis, as most studies use data from only a few years. I will estimate separate models for every year from 1975 to 2009.[8] One issue with this approach is that data availability is poor for some years; so as a general rule, I will only present and include estimates from regressions that have at least 80 observations. This threshold was chosen prior to conducting the analysis in the interest of producing estimates that cover a reasonable subset of the global population of countries. This also has the effect of producing estimates that have smaller standard errors, giving us a greater likelihood of finding significant results. In the Online Appendix, I show that the substantive findings of the paper do not appear to be particularly sensitive to this threshold and are also robust to the use of imputation instead of listwise deletion.

Table 3 summarizes the dimensions of the sensitivity analysis, which will include estimates across fourteen different public good provision measures, two measures of democracy, ninety-eight covariate sets, six lag structures, and 35 years. This yields a possible total of 576,240 models, but not all will be presented due to data coverage issues. Once low-coverage years and specifications are removed, the total settles at around 300,000 (Imai et al. 2015).

Before continuing on to the results, I should clarify that the sensitivity analysis does not purport to probe the full range of plausible assumptions and approaches, as it is impossible to do so. The analysis will consider robustness across different combinations of democracy measures, dependent variables, covariates, lag structures, and years of analysis, but I will not examine many other possibilities—non-linear or non-additive functional forms, matching approaches or pre-processing techniques, alternative structures on the error term, and so forth. I have focused on the analytical decisions that I believe to be the most critical, testing models and the neighborhood of assumptions widely used in the existing literature. This has become standard practice in other political science papers conducting systematic robustness checks (Bazzi and Blattman 2011; Dafoe 2011; Gassebner et al. 2013; Hegre and Sambanis 2006).

It is possible that democracy proves more strongly associated with public good outcomes if matching is used, if an interaction term is introduced, or if an additional covariate is included. Some readers may have alternative views as to what should be included in the vector of "fixed covariates" *I*. My results simply show that across a large range of plausible assumptions, the inference is quite fragile. Even if a specific combination of assumptions I have not tested would produce a significant estimate, this should do little to increase our overall confidence in the presence of a democratic advantage, given the abundance of evidence to the contrary.

Results

The results of the sensitivity analysis are visualized in Fig. 1a, b. The figures show the distribution of estimated *t* statistics for each of the five education and nine health

[8] In their cross-section analysis, Lake and Baum (2001) examine the two years that have the highest data coverage from each decade (1970, 1975, 1985, 1987, 1990, and 1992). Deacon (2009) presents cross-section findings from data from 1989, Moon and Dixon (1985) aggregate a quality of life index over 1970–1975, and Franco et al. (2004) rely on data from 1998.

Table 3 Dimensions of the sensitivity analysis

Dimension	Description	Total
a. Dependent variable (public good provision)	prim.pupils, sec.pupils; prim.enroll, sec.enroll, prim.completion, water, sanitation, dpt.im, measles.im, beds, physicians, lifeexpectancy, death, infmort	14
b. Independent variable (democracy measure)	Democracy (binary indicator from GWF), polity	2
c. Covariates	All models including up to four of the following concepts: (1) area; (2) popdens; (3) africa, asia, ceurope, meast, nam, sam, and scan; (4) op14 and pop14; (5) muslim and catholic); (6) aid; (7) elf.	98
d. Lag structure	Zero to five year lag structure	6
e. Year	1975–2009	35
Total possible models	a×b×c×d×e=576,240	

outcomes, for both the GWF (dark blue solid line) and Polity (red dotted line) measures of democracy. The figures also depict the line t=0, as well as the t=1.64 threshold for 5 % significance (t=−1.64 for the five "public bads": prim.pupils, sec.pupils, prim.completion,

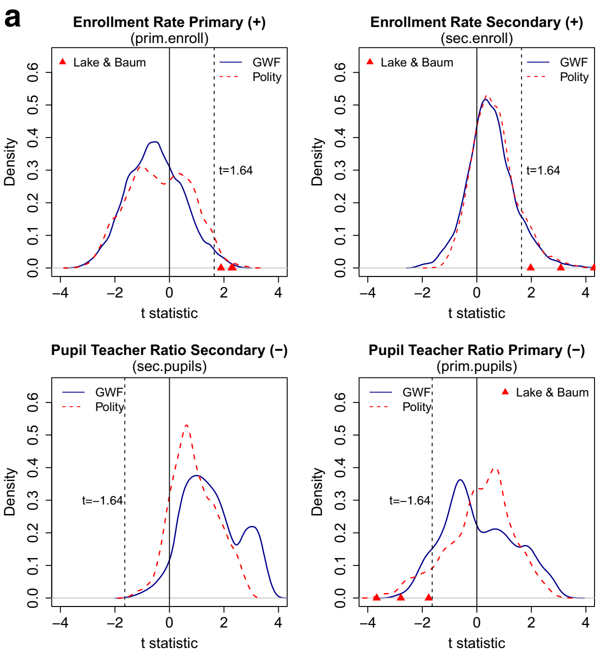

Fig. 1 Distribution of estimated t statistics for Education Outcomes. Figure summarizes distribution of estimated t statistics from cross-section global sensitivity analysis for five education public good measures from 1975 to 2009. The *straight dotted line* represents the threshold necessary to reject the null of no effect at the 0.05 level with a one-sided test. Only models with at least 80 observations are shown. All models employ OLS with robust standard errors. The predicted sign of the democratic advantage hypothesis is shown in parenthesis

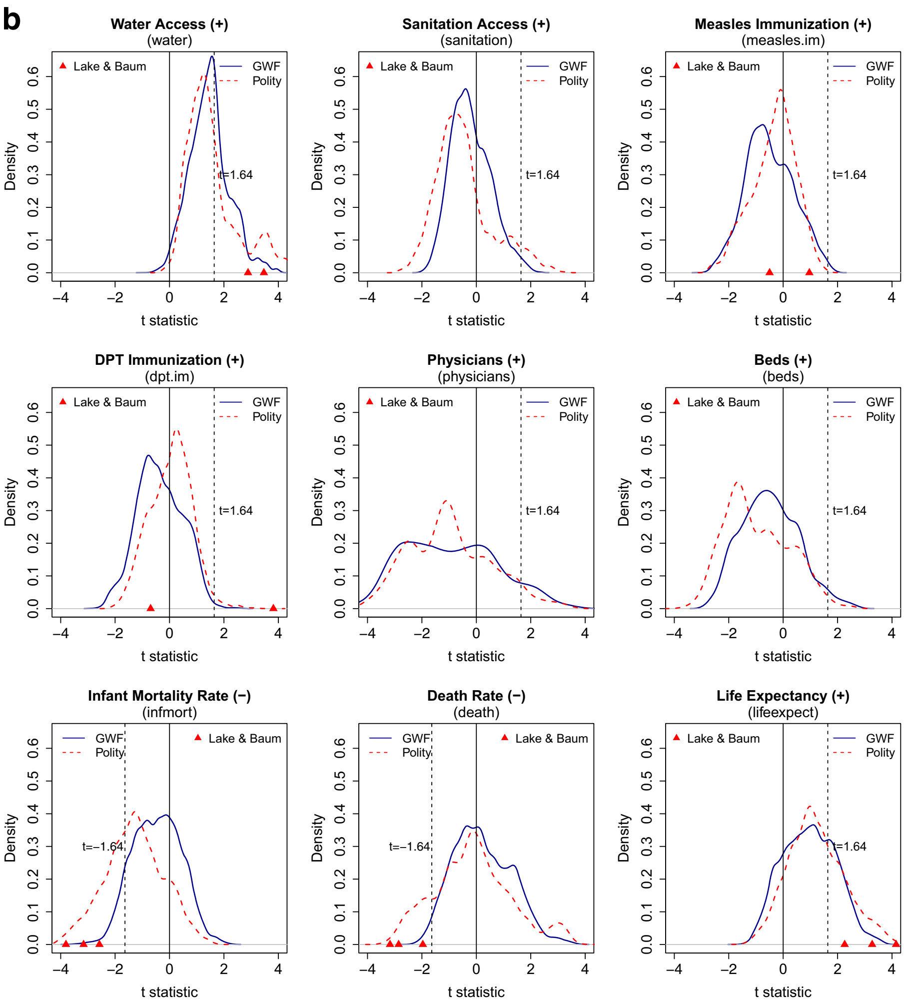

Fig. 1 (continued)

death, infmort). The democratic advantage suggests we should observe estimates far to the right for public good measures and far to the left for "public bads."

The distributions give reason to doubt the robustness of the inference. For eight of the fourteen measures—prim.pupils, sec.pupils, dpt.im, measles.im, death, sanitation, physicians, beds—the mean estimate falls on the wrong side of zero. The sec.pupils variable actually has a fairly robust association with democracy, but in the opposite direction. Authoritarian regimes appear to have fewer students per teacher in secondary schools, on average. Among the other five outcome measures, only the coefficient on water is consistently in the correct direction, and to a lesser extent, lifeexpectancy and infmort.

Figure 1a, b also depict the *t* statistics reported in Lake and Baum (2001), which tend to be highly significant and lend support to the "invisible hand of democracy" notion. These published results are misleading and lie at the extremes of the distributions.

Even for water access, the positive effect of democracy is likely much less than currently published estimates.

Tables 4 and 5 present these insights more formally using the various metrics of global sensitivity analysis developed by Sala-I-Martin and others (Gassebner et al. 2013; Hegre and Sambanis 2006; Sala-I-Martin 1997; Sturm and de Haan 2002). The columns show the CDF(0) metric, the mean coefficient estimate, the mean standard error, and average p value test. The tables also depict the fraction of estimates that reject the null of no effect at the 0.05 level with a one-sided test.

According to Sala-I-Martin's CDF(0)>0.95 test, the only dependent variable that proves to have a robust association with democracy is water. The next most robust relationship is that with lifeexpectancy, which is directionally correct 81.3 to 88.4 % of the time, depending on the democracy measure. The infant mortality rate estimates are below zero in 67.7 to 86.7 % of the specifications. As was clear in the figures, the coefficient estimates are not consistently of the correct sign for the remainder of the dependent variables. For the twenty-eight different outcome-democracy combinations presented in the tables, the CDF(0) metric exceeds 0.5 only thirteen times. This means that for more than half of the combinations, less than half of the estimates are of the correct sign.

Table 4 Global sensitivity analysis (education)

DV	Unweighted				Fraction	Models
	CDF(0)	$\overline{\beta}$	$\overline{\sigma}$	p	p<0.05	J
prim.pupils (−)						
GWF	0.536	−0.100	2.507	0.484	0.088	6494
Polity	0.413	0.015	0.184	0.532	0.091	6280
sec.pupils (−)						
GWF	0.056	2.667	1.849	0.925	0.000	3335
Polity	0.127	0.121	0.145	0.798	0.000	3285
prim.enroll (+)						
GWF	0.303	−3.099	5.126	0.727	0.021	14952
Polity	0.402	−0.137	0.346	0.654	0.031	14952
sec.enroll (+)						
GWF	0.737	2.165	4.199	0.303	0.105	12302
Polity	0.783	0.182	0.279	0.257	0.121	11848
prim.comp (+)						
GWF	0.514	0.079	3.618	0.491	0.013	4116
Polity	0.755	0.095	0.289	0.371	0.006	4116

Table summarizes results from cross-section global sensitivity analysis for five education public good measures from 1975 to 2010, for both the GWF and polity measures of democracy. The table shows the fraction of estimates that exceed zero (below zero for "public bads") mean coefficient estimate, mean standard error estimate, and average p value test. It also depicts the fraction of estimates that reject the null of no effect at the 0.05 level with a one-sided test and the total number of models. Only models with at least 80 observations are shown. All models employ OLS with robust standard errors

Table 5 Global sensitivity analysis (health)

DV	Unweighted				Fraction	Models
	CDF(0)	$\overline{\beta}$	$\overline{\sigma}$	p	$p<0.05$	J
dpt.im (+)						
GWF	0.343	−1.866	4.205	0.671	0.004	15670
Polity	0.550	0.018	0.288	0.474	0.018	15610
measles.im (+)						
GWF	0.302	−2.501	3.933	0.737	0.004	15162
Polity	0.359	−0.098	0.274	0.639	0.000	15162
water (+)						
GWF	0.985	3.992	2.631	0.064	0.636	11760
Polity	0.994	0.332	0.198	0.047	0.661	11760
infmort (−)						
GWF	0.677	−2.586	5.296	0.312	0.093	18354
Polity	0.867	−0.484	0.348	0.082	0.365	18354
death (−)						
GWF	0.456	0.114	0.880	0.551	0.019	18354
Polity	0.560	−0.016	0.058	0.389	0.156	18354
sanitation (+)						
GWF	0.359	−0.621	4.052	0.560	0.012	11466
Polity	0.243	−0.141	0.313	0.674	0.054	11466
physicians (+)						
GWF	0.310	−0.206	0.220	0.824	0.084	1352
Polity	0.269	−0.018	0.017	0.850	0.050	1352
beds (+)						
GWF	0.315	−0.269	0.594	0.674	0.029	1374
Polity	0.237	−0.040	0.041	0.837	0.013	1344
lifeexpect (+)						
GWF	0.813	1.382	1.388	0.159	0.270	18354
Polity	0.884	0.117	0.094	0.106	0.321	18354

Table summarizes results from cross-section global sensitivity analysis for nine health public good measures from 1975 to 2010, for both the GWF and polity measures of democracy. The table shows the fraction of estimates that exceed zero (below zero for "public bads"), mean coefficient estimate, mean standard error estimate, and average *p* value test. It also depicts the fraction of estimates that reject the null of no effect at the 0.05 level with a one-sided test and the total number of models. Only models with at least 80 observations are shown. All models employ OLS with robust standard errors

The democratic advantage hypothesis also fares poorly on significance metrics. For water access, the effect of democracy is significant around 63 to 66 % of the time. The one-sided average *p* value test returns values ranging from 0.047 to 0.064, which would meet conventional standards for significance. Again, lifeexpectancy and infmort are next in line, but they do not appear to consistently return significance. Democracy has a significant association with life expectancy in 27.0 to 32.1 % of specifications and an average one-sided p value ranging from 0.106 to 0.159. For the infant mortality rate, 36.5 % of estimates prove significant at the 5 % level using the Polity measure but only 9.3 % for the GWF measure. The average p value measure ranges from 0.082 to 0.312. The rest of the outcome measures return high average p values and a very low fraction of significant estimates.

The sensitivity analysis suggests that democracy has a reasonably robust positive association with water access. There is also weaker evidence that democracy is weakly associated with higher life expectancies and lower infant mortality rates. There is no evidence of a robust association in favor of the democratic advantage for the other eleven outcome measures, in sharp contrast to existing findings (Lake and Baum 2001; Deacon 2009).[9]

We might be concerned that the shape of these distributions is due to some of the decisions made in the sensitivity analysis. Recall that the figures and tables include only estimates based on at least 80 observations. Lake and Baum (2001) present some regressions with fewer than 80 observations. As Table 1 shows, the missingness problem is quite severe for many of the dependent variables—beds and physicians are missing for 63.2 and 55.6 % of country-years, respectively. Ross (2006) argues that missing data patterns—specifically the fact that high-performing autocracies are less likely to release data on public good outcomes—may also bias existing estimates, which suggests we should actually focus on estimates that have higher levels of coverage. The Online Appendix assesses the "sensitivity of the sensitivity analysis" to different choices about sample coverage. The substantive conclusions of the paper prove robust to other cutoff thresholds, as well as the use of an imputation model to address the missingness problem (Honaker et al. 2011).

The Online Appendix also explores whether the non-relationship between regime type and democracy is the result of using the GWF democracy measure. In contrast to existing binary democracy measures, the GWF dataset also delineates country-years where the territory was ruled by a provisional government, not independent, occupied by foreign troops, ruled by multiple governments, and no government at all. Other datasets, like the CGV democracy indicator, will lump these country-years (about 125 in total) into the "autocratic" category. There are also discrepancies across the two measures in terms of what constitutes democracy. Figures 4a and 4b in the Online Appendix show that these differences do not alter the substantive conclusions of the sensitivity analysis; the distributions look approximately the same regardless of whether the GWF or CGV measure is used.

Observing a couple outcomes with weakly robust associations in the correct direction is not surprising, given that the analysis investigates fourteen in total. We have also observed one with a robust association in the wrong direction (sec.pupils), and over half of the distributions of t statistics/estimates are centered on the wrong side of zero. The pooled distribution of p values also appears uniform, which is suggestive of statistical noise. There are no good reasons for suggesting that the models that do emerge as significant are "better" in some sense, nor that there are a large number of untested plausible specifications that could override the collective weight of the 300,000 specifications presented here. If we believe these fourteen dependent

[9] Lake and Baum (2001) report statistically significant results for 16 out of the 17 outcome measures they investigate. Deacon (2009) finds significance for all five of his measures.

variables are reasonably good measures of public good provision in health and education, there is nothing in the sensitivity analysis that really supports the notion of a democratic advantage.

Conclusion

There is a growing body of published empirical evidence showing a democratic advantage in public good provision (Besley and Kudamatsu 2006; Brown 1999; Boone 1996; Bueno de Mesquita et al. 2003; Deacon and Saha 2006; Deacon 2009; Franco et al. 2004; Lake and Baum 2001; Moon and Dixon 1985). This paper challenges that finding using the techniques of global sensitivity analysis. Democracy has a reasonably robust positive association with water access, and there is also weaker evidence that democracy is associated with higher life expectancies and lower infant mortality rates. However, there is no evidence of a robust association in favor of the democratic advantage for the other 11 outcome measures, in sharp contrast to existing findings. In short, once the range of plausible assumptions is considered, there are few systematic differences between democratic and authoritarian systems across a range of education and health indicators. The inference that democracy is associated with better governance on these metrics is simply “too fragile to be believed” (Leamer 1985).

This finding should by no means be taken as an endorsement of authoritarianism. Even though democracies may not be systematically better at educating their citizens and providing healthcare, there is also no evidence that authoritarian systems are systematically better, which goes against “authoritarian efficiency” arguments (Kaufman 1985). Democracies also prove superior in fostering human rights and societal openness, relationships that do survive the global sensitivity analysis. Churchill’s dictum still holds, just not because of anything to do with public good provision.

If the non-relationship between regime type and public goods is correct, it calls into question several established theories of authoritarian politics (Bueno de Mesquita et al. 2003; Deacon 2009; Lake and Baum 2001; McGuire and Olson 1996). Researchers must move beyond arguments of encompassing interest and political competition to consider the full range of differences between democratic and authoritarian regimes and how these differences connect to the nature of governance. Methodologically, this agenda may have gone as far as it can with the standard cross-national approach, and future research can do more to probe these outcomes with micro-data and stronger research designs (Harding and Stasavage 2014; Kudamatsu 2012). In moving this direction, we may well find that the key determinants of investment in teacher salaries, vaccinations to diseases, and general responsiveness to citizen demands have little to do with the presence or absence of electoral institutions, and much more to do with the hundreds of other institutions that dictate the humdrum of every day government.

Acknowledgments This manuscript was written and revised while the author received research support from Yale University, Princeton University, and the NSF Graduate Research Fellowship Program. Thanks to Allan Dafoe, Thad Dunning, Susan Rose-Ackerman, Ken Scheve, Lily Tsai, and Michael Weaver for their helpful comments on various aspects of the manuscript. Thanks also to the editorial team at SCID and the three anonymous reviewers for their support of the manuscript and constructive feedback. Any remaining errors are my own.

References

Avelino G, Brown D, Hunter W. The effects of capital mobility, trade openness, and democracy on social spending in Latin America, 1980-1999. Am J Polit Sci. 2005;49(3):625–41.
Bazzi S, Blattman C. Economic shocks and conflict: The (Absence of?) Evidence from Commodity Prices. Working Paper, 2011.
Besley T, Kudamatsu M. Health and democracy. Am Econ Rev. 2006;96:313–8.
Boone P. Politics and the effectiveness of foreign Aid. Eur Econ Rev. 1996;40:28–329.
Brown D. Reading, writing and regime type: democracy's impact on primary school enrollment. Pol Res Q. 1999;52(4):681–707.
Brown D, Hunter W. Democracy and social spending in Latin America, 1980-92. Am Polit Sci Rev. 1999;93(4):779–90.
Brown D, Hunter W. Democracy and human capital formation: education spending in Latin America, 1980 to 1997. Comp Pol Stud. 2004;37(7):842–64.
Bueno de Mesquita B, Smith A, Silverson RM, Morrow JD. The logic of political survival. Cambridge: MIT Press; 2003.
Cheibub JA, Gandhi J, Vreeland JR. Democracy and dictatorship revisited. Public Choice. 2010;143:67–101.
Coppedge M, Gerring J, Altman D, Bernhard M, Fish S, Hicken A, et al. Conceptualizing and measuring democracy: a new approach. Perspect Polit. 2011;9:247–67.
Dafoe A. Statistical critiques of the democratic peace. Am J Polit Sci. 2011;55(2):247–62.
Deacon RT. Public good provision under dictatorship and democracy. Public Choice. 2009;139:241–62.
Deacon RT, Saha S. Public good provision by Dictatorships: A survey. In: Ott AF, Cebula RJ, editors. The Companion in Public Economics: Empirical Public Economics. New York: Edward Elgar Publishing. 2006.
Franco A, Alvarez-Dardet C, Ruiz MT. Effect of democracy on health: ecological study. BMJ. 2004;329: 1421–3.
Gassebner M, Lamla MK, Vreeland JR. Extreme bounds of democracy. J Confl Resolut. 2013;57:171–97.
Geddes B, Wright J, Frantz E. New data on autocratic regimes. Working Paper. State College, PA: Penn State University Department of Political Science. 2012.
Gerber A, Malhotra N. Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. Q J Polit Sci. 2008;3:313–26.
Gerber A, Green DP, Nickerson D. Testing for publication bias in political science. Polit Anal. 2001;9:385–92.
Harding R, Stasavage D. What democracy does (and doesn't do) for basic services: school fees, school inputs, and African elections. J Polit. 2014;76(1):229–45.
Hegre H, Sambanis N. Sensitivity analysis of empirical results on civil war onset. J Confl Resolut. 2006;50: 508–35.
Honaker J, King G, Blackwell M. Amelia II: a program for missing data. J Stat Softw. 2011;45(7):1–47.
Imai K, King G, Lau O. "ls: Least Squares Regression for Continuous Dependent Variables." in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software,". 2015. http://gking.harvard.edu/zelig.
Kaufman RR. Democratic and authoritarian responses to the debt issue: Argentina, Brazil, and Mexico. Int Organ. 1985;39:473–503.
Kaufman RR, Segura-Ubiergo A. Globalization, domestic politics, and social spending in Latin America: a time-series cross-section analysis, 1973-97. World Polit. 2001;53:553–87.
Kudamatsu M. Has democratization reduced infant mortality in sub-Saharan Africa? J Eur Econ Assoc. 2012;10(6):1294–317.
Lake DA, Baum MA. The invisible hand of democracy—political control and the provision of public services. Comp Pol Stud. 2001;34:587–621.
Leamer EE. Sensitivity analyses would help. Am Econ Rev. 1985;75:308–13.
Levine R, Renelt D. A sensitivity analysis of cross-country growth regressions. Am Econ Rev. 1992;82:942–63.
McGuire MC, Olson M. The economics of autocracy and majority rule: the invisible hand and the use of force. J Econ Lit. 1996;34:72–96.
Moon BE, Dixon WJ. Politics, the state, and basic human needs: a cross-national study. Am J Polit Sci. 1985;29(4):661–94.
Mulligan CB, Gil R, Sala-i-Martin X. Do democracies have different public policies than nondemocracies? J Econ Perspect. 2004;18:75–98.
Munck GL, Verkuilen J. Conceptualizing and measuring democracy: evaluating alternative indices. Comp Pol Stud. 2002;35(1):5–34.

Przeworski A, Alvarez ME, Cheibub JA, Limongi F. Democracy and development: political institutions and well-being in the world, 1950-1990. New York: Cambridge University Press; 2000.
Ross M. Is democracy good for the poor? Am J Polit Sci. 2006;50:860–74.
Sala-I-Martin XX. I just ran two million regressions. Am Econ Rev. 1997;87:178–83.
Stasavage D. Democracy and education spending in Africa. Am J Polit Sci. 2005;49:343–58.
Sturm JE, de Haan J. How Robust is Sala-I-Martin's Robustness Analysis? Working Paper, 2002.
Wigley S, Akkoyunlu-Wigley A. The impact of regime type on health: does redistribution explain everything? World Polit. 2011;63(4):647–77.
Zweifel TD, Navia P. Democracy, dictatorship, and infant mortality. J Democr. 2000;11(2):99–114.

Rory Truex is Assistant Professor of Politics and Public Affairs, Princeton University, 440 Robertson Hall, Princeton, NJ 085444. He studies authoritarian politics with a focus on China.