ANNUAL REVIEWS

Does Democracy Matter?

John Gerring,[1] Carl Henrik Knutsen,[2] and Jonas Berge[2]

[1] Department of Government, University of Texas at Austin, Austin, Texas, USA;
email: jgerring@austin.utexas.edu

[2] Department of Political Science, University of Oslo, Oslo, Norway;
email: c.h.knutsen@stv.uio.no, jonasvb96@gmail.com

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

democracy, regime type, governance, economic policy, social policy

Abstract

Does democracy matter for normatively desirable outcomes? We survey
results from 1,100 cross-country analyses drawn from 600 journal articles
published after the year 2000. These analyses are conducted on 30 distinct
outcomes pertaining to social policy, economic policy, citizenship and hu-
man rights, military and criminal justice, and overall governance. Across
these diverse outcomes, most studies report either a positive or null rela-
tionship with democracy. However, there is evidence of threshold bias, sug-
gesting that reported findings may reflect a somewhat exaggerated image of
democracy's effects. Additionally, democratic effects are more likely to be
found for outcomes that are easily attained than for those that lie beyond
the reach of government but are often of great normative importance. We
also find that outcomes measured by subjective indicators show a stronger
positive relationship with democracy than outcomes that are measured or
proxied by more objective indicators.

INTRODUCTION

Until recently, most work in political science on the topic of regimes was devoted to democratization and democratic consolidation (see Coppedge et al. 2022). In the past few decades, attention has increasingly turned to democracy's effects. Does democracy matter?

Of course, democracy may be valued for its intrinsic virtues, such as individual freedom (civil liberty) and the ability to participate in politics (self-determination). In this light, some view political freedom as an inherent human right (Franck 1992). However, people across the world are rightly concerned with many aspects of life other than political freedom and participation. It is important to ask whether democracy produces better or worse outcomes than autocracy across other dimensions that we value.

The question has special resonance in a period when many countries seem to be turning away from democratic ideals (Lührmann et al. 2019) and when some policy makers view the authoritarian "Chinese model" as a viable alternative for promoting good governance and development (Li 2015). In academia, influential voices have cast doubt on electoral accountability as a mechanism of good governance (Achen & Bartels 2017) or proposed that other institutional features such as strong state institutions are more relevant for ensuring good governance and development (Fukuyama 2014, Rothstein 2011).

Past reviews have concluded on an ambivalent note, suggesting that democracy is relatively inconsequential for many policies and policy outcomes (Carbone 2009, Doorenspleet 2019, Mulligan et al. 2004). However, these reviews were not very comprehensive and several were undertaken some time ago, prior to the recent avalanche of work on democracy's effects.

Granted, several recent meta-analyses and literature reviews focus on the democracy–economic growth relationship (Colagrossi et al. 2020, Knutsen 2021), suggesting that democracy may have a more beneficial impact than was indicated by earlier overviews (e.g., Przeworski & Limongi 1993). However, comprehensive, up-to-date assessments of the literature pertaining to most other governance outcomes are missing.

In this article, we assess more than 1,100 cross-country analyses drawn from more than 600 journal articles published after the year 2000. Our survey covers a vast range of governance outcomes pertaining to social policy, economic policy, citizenship and human rights, military and criminal justice, and overall governance.

If this literature is to be believed, the agnostic conclusion of previous reviews no longer seems plausible. But neither does the optimistic assertion that democracy always produces good governance and benevolent outcomes (Halperin et al. 2009). Rather, there seems to be ample variation. Positive effects are usually reported for outcomes pertaining to human rights, trade, transparency, quality of government, and health/human development. Weak or null results predominate in analyses focused on social transfer programs, inflation, inequality, and public spending. Relatively few analyses suggest that democracy worsens the quality of governance, and even fewer find a strong negative effect. Based on this review of the literature, one might conclude that democracy's impact on governance outcomes is either positive or null.

Although the main purpose of this review is to elucidate published findings, we also address the likely validity of this body of research. We find evidence of threshold bias (around the 0.05 significance level) and we cannot dismiss the possibility of normative bias (toward "desirable" results; i.e., that democracy is associated with normatively desirable outcomes). We hope that this review makes a substantive contribution and also points the way to a new approach to conducting systematic reviews of areas where there are multiple, distinct outcomes (rendering a standard meta-analysis problematic). To encourage future work that takes a broad-angle view of democracy's effects, we refer readers to our **Supplemental Database**.

We begin by laying out the obstacles and payoffs attendant on examining democracy as a causal factor. Next, we describe the extensive database of published studies that we have collected. Thereafter, we summarize the findings contained in this literature. Finally, we examine possible biases contained within those findings.

DEMOCRACY AS A CAUSE

Democracy is a manipulable treatment only in the loosest sense. Accordingly, when someone asserts, "Democracy causes Y," it is unclear what aspect of democracy is affecting the outcome or if all aspects of this composite concept are contributing equally. Nor is it clear what the background (ceteris paribus) conditions are.

In light of these difficulties, it might seem advisable to re-specify the research question in a manner that comes closer to the potential-outcomes vision of causality. For example, one might focus on specific components of democracy (e.g., multi-party elections, universal suffrage) or types of democracy (e.g., presidentialism, federalism).

While a more narrowly defined research agenda holds promise, one must appreciate that smaller interventions usually have smaller effects, making it more difficult to distinguish treatment effects from background noise in the limited samples available. Moreover, the various elements that compose the concept of democracy interact with each other, making it difficult to analyze them independently. Imagine a scenario in which 20 relevant elements are associated with democracy, and the author chooses to focus on just one that is of special theoretical interest. In this scenario, one would be obliged to measure and condition on the other 19 to rule out potential confounders—a daunting task.

The fact that political institutions interact with each other also makes it challenging to aggregate up our understanding of the parts of democracy into an understanding of the whole. Political systems exhibit strong holistic properties. During periods of regime change, many of their component elements change together.

Thus, despite its ambiguities, the macro-level question of regimes remains indispensable. We need to know whether being governed by democratic rules matters for outcomes that we care about. We believe that it is possible to bring democracy into the potential-outcomes framework so long as these ambiguities are recognized and claims are suitably qualified by caveats. Specifically, we view democracy as a composite variable that can, in some instances, be considered as an as-if-random treatment. One such occasion is foreign intervention—either prodemocratic (e.g., the Allied occupation and reconstruction of Germany and Japan) or antidemocratic (e.g., the US-engineered coups in Chile and Iran). Another occasion is when leaders accidentally inaugurate a democratic transition (e.g., Gorbachev in the USSR) (Treisman 2020).

Unfortunately, these events are rare. They may also display heterogeneous effects in different contexts. Consequently, it is difficult to arrive at generalizable causal effects by examining isolated as-if-random treatments. Accordingly, we are consigned to examine this question with data that are largely observational and mostly at the country level.

This review is limited to large-n cross-national studies. We realize that this focus eliminates from consideration a good deal of information, especially where studies enlist both qualitative and quantitative evidence and where evidence is drawn from across different units of analysis (e.g., subnational regions and individuals). Unfortunately, it is infeasible to integrate all research design components into a single review.

We do not undertake a traditional meta-analysis, as has been done for democracy and growth (Colagrossi et al. 2020, Doucouliagos & Ulubaşoğlu 2008). Given the wide range of outcomes under consideration it is difficult to standardize and meaningfully compare coefficient estimates using the standard meta-analysis tools.

Despite these difficulties, we think something can be learned from systematically analyzing the body of work conducted on our topic. After all, there is *some* independence across studies, as well as across analyses within a given study. This can be seen in the immense diversity of approaches taken to our subject, surveyed in the next section. One is struck by the sheer number of decisions authors must make in order to arrive at a benchmark model. These are consequential decisions, as we shall see. Although one might regard some approaches as "better" in the sense of setting a higher standard of inference, it is hard to determine—even for a given input and output—what the best methodological approach might be.

What is the best approach depends to a certain extent on the nature of the data available. Some outcomes can be measured over a long time period and with considerable year-to-year variance; other outcomes are sluggish or can be measured only for recent years. Specification strategies also depend on the outcome under investigation. But all decisions depend on authorial judgments, and these judgments are not easy to classify as correct or incorrect. The many degrees of freedom available to scholars working with observational data enhance our sense that we might learn something by collating and comparing those decisions.

To be sure, these expert judgments are not entirely independent. Authors read each other's work and are influenced by input from reviewers and editors, as well as general intellectual currents. Nonetheless, given the many choices and lack of absolute standards, the body of work published on a subject constitutes a series of data points that may provide a more reliable estimate of democracy's effects on an outcome than any single study observed in isolation.

A DATABASE OF STUDIES

In order to be considered in our survey, a study must explicitly focus on the causal role of democracy. We exclude studies where democracy forms a background variable of no apparent theoretical interest, and where results for the democracy variable are not explicitly interpreted. Such studies may be misleading as the construction of empirical tests has another purpose, which may not be well-devised for testing the role of regimes.

Studies are eligible if they include at least 60 countries. This rather high threshold is intended to exclude studies that focus on particular world regions (e.g., Africa or the OECD). Doing so ensures that we maintain a focus on general patterns and that there is some degree of comparability across research designs. We also exclude studies that focus only on interaction effects.

We acknowledge that relationships of interest may operate differently in different parts of the world and in different contexts, being responsive to level of development, size, culture, colonial history, and other background factors. In a global data set, this sort of heterogeneity often attenuates estimates of the average treatment effect. However, our theoretical interest is in effects that are broadly generalizable. Likewise, we do not attempt to track nonmonotonic relationships. It is sometimes claimed, for example, that democracy's impact on corruption is curvilinear, increasing and then decreasing (McMann et al. 2020). In situations like this, we examine only the estimated monotonic effect (if any).

Outcomes of interest extend to most policies and policy outcomes associated with the wide-ranging concept of governance. We are interested in policy effort (e.g., the presence or absence of particular policies or expenditures devoted to them), political processes (e.g., corruption), and outcomes that might be regarded as the end-product of policies (e.g., crime rates or infant mortality rates). We embrace 30 types of outcomes (see **Table 1** for details) grouped under five broad headings: social policy, economic policy, citizenship and human rights, military and criminal justice, and overall governance. We exclude outcomes related to societal conflict and political instability—social unrest, riots, demonstrations, civil wars, other wars, genocide, ethnic

Table 1 Outcomes, indicators, and normative implications in five policy areas: social policy, economic policy, citizenship and human rights, military and criminal justice, and overall governance[a]

Outcome category	Desirable outcomes, by indicator
Social policy	
Human development, health	Low mortality, increased health spending, high immunization rates
Food, nutrition	High calorie consumption
Education	High educational attainment, high education spending, high literacy rates
Environment	Low emissions, low deforestation rates, sustainable policies, more protected areas
Demography	Low fertility, high urbanization
Disaster preparedness	Few disaster deaths, low disaster costs
Social transfer programs	High coverage of pension systems, effective social security programs, more social spending
Inequality	Low income inequality, high fiscal redistribution
Public spending	*High government spending, high government consumption*
Economic policy	
Agricultural subsidies	*High agricultural subsidies*
Trade	More trade, trade openness
Technological change	High total factor production growth, internet and phone diffusion, innovation, quality of E-government
Regulation	Secure property rights, economic freedom, deregulatory reforms
Fuel subsidies	*Reduced fuel subsidies*
Foreign direct investment	High foreign direct investment
Investment	High investment, financial development
Inflation	Low inflation
Monetary policy	*Central bank independence reform*, *floating exchange rates*, money supply growth, *foreign exchange reserves*
Fiscal policy	Low debt, *high tax revenue*, low current account deficits, high credit access
Wages, employment	High wages, high employment
Infrastructure	High-quality infrastructure, water and electricity access, manufacturing
Tourism	*High tourism, tourism competitiveness*
Growth	Economic growth, low growth volatility
Citizenship and human rights	
Human rights	Gender equality, civil rights, physical integrity rights
Migration	*High immigration, low emigration*
Military and criminal justice	
Military spending	*Low military spending*
Criminal justice	Low crime rates, abolition of death penalty, low prison population
Overall governance	
Transparency	Fiscal transparency, national transparency, freedom of information laws
Corruption	Low corruption
Quality of government	State capacity, effective government, rule of law, perception of good governance

[a]Outcomes whose normative import is questionable are printed in italics. For presentational purposes, relatively similar indicators are grouped together. For example, "growth volatility" encapsulates three indicators: growth volatility, growth reversals, and growth accelerations.

cleansing, coups, assassinations, leader turnover, purges, government crises, terrorism, revolution, and constitutional change—which we regard as a separate area of study.

Our review is limited to articles published in peer-reviewed English-language journals. Books are excluded, as the format tends to be reserved for more discursive material (or reprints of empirical analyses already published in journals). We leave aside working papers and other unpublished materials. Although this approach is bound to reflect existing publication biases, we see no easy way around the problem. Unpublished material is often unavailable, and available papers are in any case likely to be constructed for publication and thus equally susceptible to "fishing."

Among published articles, we limit our survey to studies published in the twenty-first century. This cutoff date is arbitrary, to be sure, but it seems reasonable given the explosion of cross-national data in comparative politics over the past two decades. Additionally, we expect that recent studies are likely to employ more sophisticated analyses, a product of improved graduate training, methodological and software developments, and increased attention to causal inference.

To identify as many studies as possible that fit these selection criteria, we adopt a multipronged strategy. Several search engines are employed, including Google Scholar, Microsoft Academic, ScienceDirect, and Oria (to search in the University of Oslo library). We employed a wide range of search terms to capture studies on various outcomes published in different disciplines. For each (recent) study, we also carefully perused the references.

No protocol is perfect, and our procedure may have missed some studies that fit our selection criteria. But we are fairly certain not to have missed any influential studies, as these are cited by numerous other papers and appear at the top of search results. Accordingly, the resulting sample should offer a reasonable representation of recent scholarship on these topics.

A plurality of studies (34%) involve only one outcome, measured by one indicator. However, some studies employ several indicators for a single, overarching concept. For example, Coccia (2020) measures "innovation" with counts of internet users, personal computers, cellphone subscribers, and patents per capita. Other studies include multiple outcomes. For example, Giavazzi & Tabellini (2005) explore corruption, economic growth, fiscal policy, investment, monetary policy, and quality of government. To capture all relevant variability across studies, outcomes, and indicators, we code them separately. Each row in our database consists of a unique analysis—formed by the combination of (*a*) study, (*b*) outcome, and (*c*) indicator. A change in any of these elements warrants a separate analysis. (Minor changes in measurement, such as log-transforming the original indicator or using a similar measure from a different data set, do not count as changes in indicators but rather as robustness tests.)

The resulting database includes 607 studies (from 225 distinct journals), 30 outcomes, 212 indicators, and 1,181 analyses (see **Supplemental Database**). Selected features of this database are summarized in **Table 2**.

For each study, we record a complete bibliographic reference, URL, authors, publication year, journal, journal discipline, and abstract. Most analyses, and most studies, are drawn from political science (42%) and economics (37%), with a much smaller number from sociology (3%) and the remainder categorized as interdisciplinary (19%).

For each analysis, we identify one statistical model as the benchmark. This is often the first specification presented in the article or is identifiable by the author's explicit statement of a main, benchmark, or baseline model. When left implicit, we make a judgment call based on how authors present their results and interpret their findings.

Several features pertaining to the benchmark model are recorded. First, we note the chosen measure of democracy and its source, and whether that measure is binary or not. Among democracy data sets, Polity is by far the most common, followed by Freedom House and Democracy-Dictatorship, as noted in **Table 2**.

Table 2   Analyses of democracy's effects

Discipline (categorized by journal): number of analyses (number of studies)	Number of analyses (number of studies)	Policy areas	Number of analyses
Economics	432 (232)	Social policy	329
Political science	491 (240)	Economic policy	483
Sociology	36 (21)	Citizenship, human rights	96
Other (mainly interdisciplinary)	222 (114)	Military, criminal justice	75
Democracy data set (nonexclusive categories)	**Number of analyses**	Governance overall	196
Boix, Miller, Rosato	30	**Attainability of outcome measure**	**Percent of analyses**
Democracy-Dictatorship	93	Hard	44
Database of Political Institutions	16	Intermediate	13
Economist Intelligence Unit	21	Easy	43
Freedom House	222	**Operationalization of outcome measure**	**Percent of analyses**
Papaioannou, Siourounis	12	Subjective assessment	16
Polity	719	**Sample**	**Mean**
Unified Democracy Scores	4	Countries	115
Vanhanen	13	Years	27
Varieties of Democracy	18	**Benchmark model**	**Percent of analyses**
World Governance Indicators	4	Cross-section	32
Other	56	Ordinary least squares	66
Democracy scale	**Percent of analyses**	Country fixed effects	30
Binary	24	Year fixed effects	37
		Lagged dependent variable	20
		Any error correction	66
		Extensive robustness checks	34

Second, we register the outcome measure and its data source. Each measure is coded according to whether it involves a substantial degree of subjective assessment, e.g., country-expert perceptions of corruption.

We also code outcome measures according to their attainability on a three-point scale. An attainable measure is one that is in the capacity of a government to achieve, given sufficient political will. Examples of attainable measures include changing the terms of a statute or government order (e.g., abolishing capital punishment) and devoting money to the policy (expenditures). A difficult-to-attain measure is one like growth or infant mortality that is affected by many factors outside the immediate control of government. An intermediate category is used for measures that fall somewhere in the middle. (When a second author recoded 210 observations, 82% of original scores were replicated, a strong signal of intercoder reliability.)

For the benchmark model, we note several features of the sample including the first year, last year, number of countries, and number of total observations. We also note features of the research design including data structure (e.g., panel, country-year), country fixed effects, year fixed effects, inclusion of other temporal controls, lag structure (for how many years is democracy measured prior to the outcome?), a lagged dependent variable, estimator, type of standard errors, and a covariate list.

With respect to the finding—democracy's estimated impact on the chosen outcome—we code the direction of the relationship (positive or negative), statistical significance level, and *t*-value. In
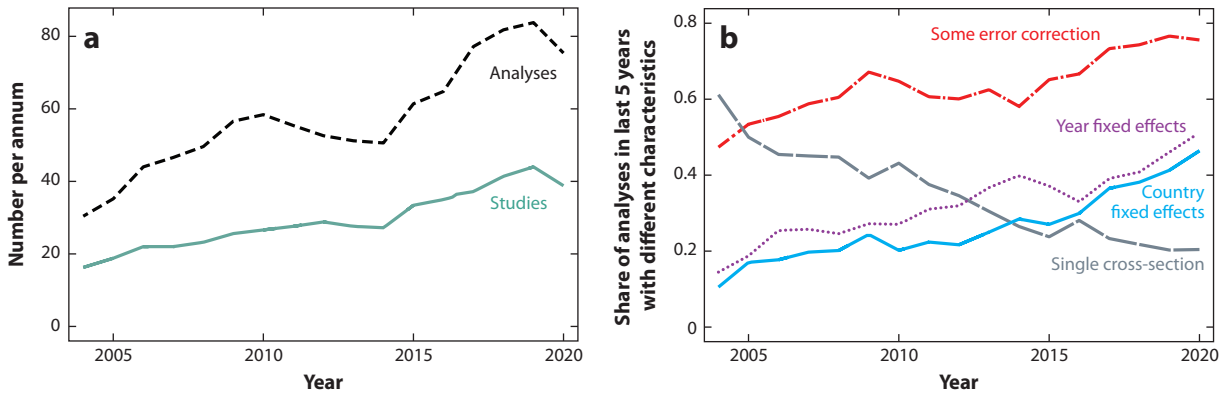
Figure 1

Time trends in the study of democracy's impact on governance outcomes. (*a*) Number of studies and analyses per annum recorded in our database of published cross-country studies, smoothed across a moving five-year period. (*b*) Share of analyses over a five-year period that display particular methodological characteristics: single cross-section (a single year or range of years collapsed into one point estimate), country fixed effects, year fixed effects, and some error correction (errors are not classical). Authors' coding.

evaluating democracy's effect on governance, we rely primarily on *t*-values, as reported in published studies or calculated from reported coefficients and standard errors. The *t*-value is not an effect, strictly speaking; it measures the estimated effect relative to the standard error. However, the range of outcomes and measures of democracy incorporated here makes it tricky to standardize effect sizes. Another benefit of the *t*-value is that it incorporates uncertainty into a single statistic, furthering our goal of making concise comparisons across myriad studies. Finally, we code a binary measure capturing the number of robustness tests presented (few/more).

Across the observed period, there are several interesting trends. **Figure 1*a*** shows the total number of studies and analyses, reported as annualized averages across a five-year window. One can see a clear upward trend across the past two decades.[1] Trends by discipline (e.g., economics or political science) are very similar (**Supplemental Appendix A**). Democracy's impact on governance is clearly a topic of growing interest.

Figure 1*b* shows evidence of increasing methodological sophistication. In the early 2000s, approximately 60% of analyses involved a single cross-section of countries, whereas only about 20% did so around 2020. Likewise, the share of analyses controlling for country and year fixed effects increased from less than 20% to around 50%, and the share employing some form of error correction (e.g., to account for temporal autocorrelation) increased from below 50% to almost 80%.

Other notable trends are displayed in **Supplemental Appendix A**. For instance, the share of analyses accompanied by extensive robustness checks has increased markedly. Moreover, researchers typically draw on more extensive samples today than 20 years ago; the availability of data has increased in both country coverage and time series coverage. Thus, according to various measures of methodological adequacy, there has been notable progress in this growing field of research.

---

[1]The attenuation at the very end reflects the outlier-year 2015 (110 analyses) being excluded from the 5-year period, and the relatively low count for 2020 (63), as coding for several outcomes was done prior to the end of 2020.

DOES DEMOCRACY IMPROVE GOVERNANCE?

Our main objective is to evaluate what the scholarly literature says about democracy's impact on governance. To examine this question, we must first code each outcome indicator according to its normative import. Is it widely regarded as good for society?

For most outcomes the answer is fairly self-evident. Most people will agree that increases in infant mortality are bad. So, if an analysis shows that democracy is correlated with lower infant mortality, we count this as a "positive" effect, reflecting that democracy is correlated with the normatively desirable outcome—with the usual ceteris paribus caveat.

Not all outcome indicators are as easy to classify. To get a sense of our judgments, **Table 1** provides a comprehensive list of outcomes and types of indicators (categorized broadly) and shows how we classify their normative implications. Judgments that seem potentially problematic—over which reasonable people might disagree (about 15% of the analyses)—are printed in italics.

In a few cases, the import of outcomes varies by context. For example, high fertility, low tax revenue, and urban bias are often seen as problems in the developing world while low fertility, high taxes, and agricultural subsidies are often viewed as problematic in the developed world. Because most countries analyzed in the reviewed studies are developing countries, we code these outcomes according to their impact in that setting.

These judgments form the basis of **Figure 2**, where we plot normatively adjusted t-values for all analyses. A normatively adjusted t-value is signed positively if democracy is correlated with a desirable outcome and negatively if correlated with an undesirable outcome. In **Supplemental Appendix A**, we present a figure that excludes analyses with difficult-to-classify outcomes (italicized in **Table 1**). The resulting density plot is very similar to **Figure 2**.

The distribution in **Figure 2** is clearly right-skewed; only 22% of t-values are negatively signed. One plausible interpretation is that democracy has desirable effects on most of the governance
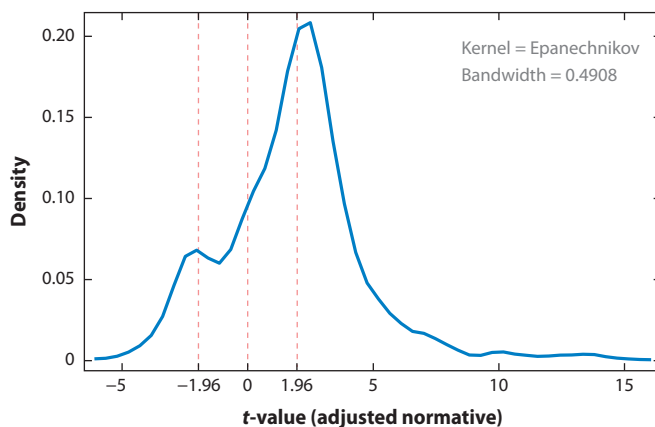

Figure 2

Kernel density plot of t-values, normatively adjusted. The figure displays t-values for the democracy coefficient in 985 cross-country analyses of democracy's impact on governance outcomes, coded as positive ($t > 0$) if democracy is correlated with an outcome that is deemed to be good for society (see **Table 1**). Extreme values (lowest and highest 1%) are removed to improve visualization. Imprecisely calculated t-values, where the coefficient and the standard error are listed in the original study with single digits, are removed to avoid artificial lumping. The Epanechnikov kernel function is used to construct the plot. The red dotted lines mark t-values of −1.96, 0, and 1.96. Authors' coding.

outcomes that scholars have chosen to examine. Of course, this is not the only interpretation, and we discuss potential biases later. For the moment, we focus on what the literature claims to find.

Heterogeneous Effects

Because our review encompasses a wide range of outcomes, it seems likely that the curve illustrated in **Figure 2** hides a good deal of heterogeneity across policy areas. To investigate this possibility, **Table 3** disaggregates our subject into the 30 outcome categories listed in **Table 1**. For each category, we show the median, mean, and standard deviation of the normatively adjusted t-values in the leftmost columns, as well as the number of analyses and studies.

In the middle columns, we classify analyses into three categories according to their (normatively adjusted) t-values, depending on whether they attain negative or positive t-values exceeding the standard 1.96 threshold. We recognize the danger of relying too heavily on arbitrary thresholds, especially given evidence of publication bias (discussed below). Even so, published studies take these thresholds seriously, and so do many readers. Accordingly, one may crudely classify analyses according to whether democracy has bad, null, or good effects on governance.

In the rightmost columns, we report the number of analyses and mean normatively adjusted t-values when all normatively ambiguous outcomes (181 analyses, italicized in **Table 1**) are excluded. After those exclusions, results in some outcome categories have only a few or no analyses left. Yet, for most, outcomes results are similar.

When consulting **Table 3**, readers should keep a close eye on columns 3–4, measuring t-value variability and total number of analyses. Some categories, such as food/nutrition and trade, are subject to enormous variability. Further inspection shows that this is largely driven by one study in each category with extremely large t-values (48.9 and 71.4, respectively) and no error correction.

Other categories, such as fuel subsidies and wages/employment, encompass just a few analyses. Our statistics are subject to the stochastic error inherent in all small samples. Indicatively, had we extended our sample back one additional year, two analyses from Rodrik's (1999) landmark study finding positive relationships with democracy would have been added to the three recorded ones, and the average t-value for wages/employment would have been $+1.6$ instead of -1.6. We caution against drawing strong conclusions for any outcome represented by fewer than 10 analyses.

Based on all the information contained in **Table 3**, the case for democracy as a benevolent factor in governance appears quite strong for the outcomes of human rights, transparency, corruption, human development/health, trade, quality of government, technological change, and foreign direct investment.

There is similar evidence for a benevolent effect of democracy on military spending, agricultural subsidies, and migration, although these conclusions hinge on accepting our judgments about what is regarded as desirable for normatively ambiguous outcomes—and are hence excluded in the right-most columns of **Table 3**.

The case for a benevolent democratic effect is pretty strong for the outcome of criminal justice, as well as five social and economic policy outcomes—regulation, education, environment, growth, and monetary policy. Median t-values are close to 2 for these outcomes.

Although relatively few studies suggest that democracy produces undesirable consequences, results from the surveyed studies are inconclusive for the outcomes of fiscal policy, infrastructure and industry, investment, social transfer programs, and disaster preparedness. For inequality, inflation, and public spending, there is no case for democracy producing desirable outcomes, according to the metrics in **Table 3**.

In summary, our review produces evidence that democracy is positively related to normatively desirable outcomes in many policy areas. Remarkably, we do not find any outcome category

Table 3   Heterogeneous democratic effects[a]

Outcomes	t (median)	t (mean)	t (SD)	Analyses (n)	Studies (n)	"Bad" t ≤ −1.96 (%)	"Null" 1.96 < t > 1.96 (%)	"Good" t ≥ 1.96 (%)	Unambiguous outcomes t (median)	Unambiguous outcomes Analyses (n)
Human rights	3.5	4.8	4.7	81	44	5	21	74	3.4	76
Military spending	2.9	4.6	5.3	27	23	4	19	78	NA	0
Transparency	2.5	3.3	3.1	36	14	3	31	67	2.5	36
Corruption	2.5	2.7	4.3	54	51	9	22	69	2.5	54
Human development, health	2.5	2.7	3.7	79	41	8	28	65	2.5	79
Trade	2.4	3.7	11.2	43	30	12	14	74	2.4	41
Quality of government	2.4	2.6	5.3	83	44	7	28	65	2.6	63
Technological change	2.2	2.3	3.1	52	33	6	39	56	2.2	51
Tourism	2.2	2.3	0.7	6	4	0	50	50	NA	0
Agricultural subsidies	2.2	2.0	1.7	12	10	8	17	75	NA	0
Regulation	2.2	1.3	2.6	48	27	10	38	52	2.4	46
Migration	2.1	3.0	4.4	11	8	9	27	63	NA	0
Foreign direct investment	2.1	2.2	1.4	22	21	5	36	59	2.1	22
Education	2.1	1.8	1.9	52	33	6	40	54	2.0	50
Criminal justice	2.1	1.6	2.4	45	20	13	29	58	2.0	42
Environment	2.0	1.6	2.5	77	37	13	38	49	2.0	76
Growth	2.0	1.4	4.7	83	74	8	37	54	2.0	82
Monetary policy	1.8	1.7	4.4	21	20	19	33	48	0.7	6
Fiscal policy	1.8	1.0	2.4	47	31	17	40	43	1.0	14
Infrastructure and industry	1.5	0.9	2.3	44	18	11	45	43	1.3	38
Investment	1.3	0.9	3.0	25	19	16	48	36	1.3	25
Social transfer programs	1.1	1.0	2.2	36	14	17	56	28	1.1	33
Disaster preparedness	0.9	1.1	2.6	36	13	6	67	28	0.9	36
Food, nutrition	0.8	8.2	20.0	6	5	17	50	33	−0.2	5
Demography	0.8	2.4	7.1	5	5	40	20	40	−2.3	3
Inflation	0.7	0.2	2.5	20	19	25	45	30	0.7	20
Inequality	0.6	0.9	2.2	37	25	19	49	32	0.6	36
Public spending	0.5	0.0	2.2	23	20	26	57	17	0.6	2
Fuel subsidies	−0.1	−0.1	2.1	6	6	33	50	17	NA	0
Wages, employment	−1.9	−1.6	0.6	3	2	33	67	0	−1.9	3
Across all analyses	*2.0*	*2.1*	*4.5*	*1120*	*577*	*11*	*35*	*54*	*2.1*	*939*

Abbreviations: NA, not applicable; SD, standard deviation.

[a]All t-values are normatively adjusted, with positive values indicating that democracy is correlated with a desirable outcome. Imprecisely calculated t-values are included in this table. Sixty analyses were dropped as both t-values and information for calculating t-values were unattainable, e.g., because they only report coefficients and significance stars. The number of studies summed across the outcome categories exceeds the total number of studies since several studies include analyses falling into multiple categories.

(with more than a few analyses) for which democracy produces a strong negative (undesirable) effect.

ARE PUBLISHED STUDIES TRUE?

To say that scholarly studies have become more numerous and more sophisticated (see **Figure 1**) does not necessarily mean they are getting closer to the truth. Thus far, we have approached the question of democracy's effects as a matter of assertion, attempting to determine what the literature says. Now, we must consider whether published results are truthful, on the whole, or whether there might be systematic biases. We explore two sorts of biases, which we characterize as threshold bias and prodemocracy bias.

Threshold Bias

Previous analysis suggests that studies are more likely to be published if they have statistically significant results (Brodeur et al. 2020, Gerber & Malhotra 2008). Bias arises if authors discard work that does not fit the bill, if they fish for a particular specification that renders the desired result (aka p hacking), or editors and reviewers refuse to publish papers that fail to produce that result. We refer to this as threshold bias, as the term is more specific and operational than the usual moniker, publication bias. (Typically, publication bias is inferred from threshold bias, although the former is not directly observed.)

To assess the matter from the record of published studies, we plot the frequency distribution of t-values. (Other tests, such as the well-known funnel test, require both coefficients and standard errors and are designed for situations in which inputs and outputs are identical, or nearly so.) If we find heaping just to the right of an established significance threshold, we may infer that published studies distort the actual distribution of findings one would expect if scholars, reviewers, and journal editors were unconcerned with beating the threshold.

Figure 3 shows that this indeed is the case, suggesting that $t = 1.96$ (corresponding to $p = 0.05$) is the critical value that many scholars and journals seek to achieve. We find that 11% of the analyses achieve absolute t-values in the fairly narrow interval 1.96–2.26. (There is no equivalent heaping at t-/p-values of 1.65/0.10 or 2.58/0.01, although the intervals following these points are also densely populated.)

The bulge at 1.96 is concerning, although it is worth noting that the sample also contains a considerable number of results that are insignificant even at the 10% level (29%) or have t-values above 3.0 (32%). One can surmise that many studies with t-values just above the 1.96 threshold probably should be accorded a slightly lower t-value. Of course, they would still be positive, and quite a distance from zero. As such, one might argue that they distort the true distribution only slightly.

The bigger concern arises from the studies that we cannot observe because the authors were unable or unwilling to finagle a statistically significant result and therefore shelved the project or met with rejection in the review process. In either case, findings lie unobserved in the proverbial file drawer. It is possible that, were all of these studies to be brought to light, the kernel density plot of absolute t-values in **Figure 2** would shift much further toward the null.

Prodemocracy Bias

A second and arguably more serious bias occurs if an entire field is oriented toward a particular finding. It is our sense that most people who study the effects of regimes have a strong normative commitment to democracy. Some scholars undoubtedly see their work as advancing the cause of
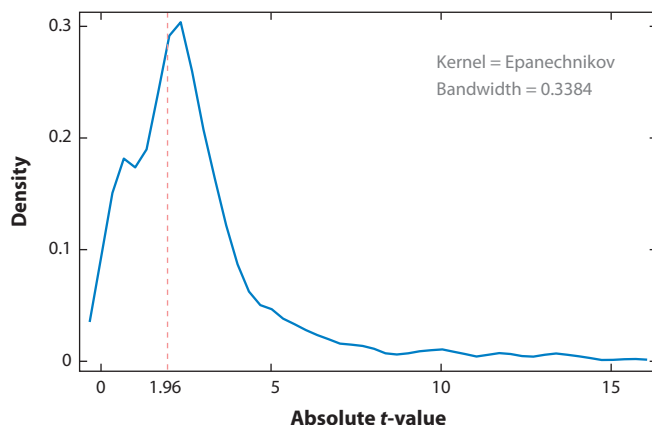
Figure 3

Kernel density plot of *t*-values, absolute value. The figure displays *t*-values for the democracy coefficient in 985 cross-country analyses of democracy's impact on governance outcomes, transformed by the absolute value (so all values are positive). Accordingly, this figure does not indicate anything about the direction of the results but only statistical significance. Extreme values (highest 1%) are removed to improve visualization. Imprecisely calculated *t*-values, where the coefficient and the standard error are both listed as single digits, are removed to avoid artificial lumping. The Epanechnikov kernel function is used to construct the plot. The red dotted line marks an absolute *t*-value of 1.96. Authors' coding.

democracy in the world. As such, they may be especially motivated to show that democracy has good effects. The same motivations are likely prevalent among journal editors and reviewers.

Of course, it is also possible that scholars have a stronger commitment to the profession and to professional norms than to advancing the cause of democracy. Or they may see the latter as contingent on the former—false or exaggerated claims probably do not advance the cause of democracy over the long haul. From this perspective, ensuring the independence and integrity of science is the best way for academics to safeguard democratic values.

There is no directly observable feature that would inform us about how strong prodemocracy bias might be or whether it exists at all. We can take some satisfaction from noting that threshold bias seems to exist at both ends of the spectrum, as shown in **Figure 2**. Yet, the heaping of *t*-values is more substantial next to the +1.96 threshold than the −1.96 threshold.

Unraveling the Data-Generating Process

To learn more about these potential biases, we interrogate our sample in a series of regressions shown in **Table 4**. None of these results should be regarded as definitive. Nonetheless, they offer clues about the data-generating process behind our set of published studies.

To investigate threshold bias, we assign a code of 1 if an analysis carries an absolute *t*-value ≥ 1.96, and 0 otherwise. This binary variable forms the outcome in Models 1–2. To assess prodemocracy bias, we adopt the (continuous) normatively adjusted *t*-values shown in **Figure 2** as the outcome in Models 3–4. Hence, Models 1–2 test which factors might be conducive to statistically significant results, while Models 3–4 test which factors are conducive to results that reflect favorably on democracy's impact on governance.

For each outcome, we offer two specifications. The first includes only the right-side variables of theoretical interest. The second adds dummies representing the 30 outcome areas, as listed in **Table 3**.

Table 4 Factors affecting significance and direction of democracy coefficient

| | Statistical significance | | | | Prodemocracy | | | |
| | Significant at 5% | | | | Normatively adjusted t-value | | | |
Dependent variable	1		2		3		4	
Discipline (reference category = Economics)								
Interdisciplinary	0.090[a]	(1.76)	0.084	(1.64)	0.081	(0.17)	–0.126	(–0.29)
Political science	0.105[b]	(2.53)	0.076[a]	(1.80)	0.184	(0.48)	–0.204	(–0.48)
Sociology	0.187[b]	(2.57)	0.202[b]	(2.22)	0.144	(0.19)	–0.163	(–0.21)
Journal								
CiteScore (ln)	–0.084[b]	(–2.24)	–0.088[b]	(–2.55)	– 0.218	(–0.72)	–0.286	(–0.92)
Time-trend								
Year of publication	–0.073[c]	(–3.13)	–0.065[c]	(–3.06)	–0.313	(–1.46)	–0.282	(–1.23)
Breadth								
Analyses per study (ln)	–0.013[c]	(–3.81)	–0.014[c]	(–4.05)	–0.022	(–0.70)	–0.009	(–0.29)
Design								
Countries (n)	0.002[c]	(3.37)	0.001[c]	(2.59)	0.016[c]	(2.69)	0.009	(1.58)
Time series (years)	0.001[a]	(1.79)	0.001[b]	(2.20)	0.001	(0.28)	0.002	(0.42)
Single cross section	0.045	(0.89)	0.031	(0.63)	–0.578	(–1.21)	–0.458	(–1.00)
Country fixed effects	0.009	(0.18)	0.020	(0.43)	–0.327	(–0.81)	–0.259	(–0.61)
Year fixed effects	0.023	(0.48)	0.004	(0.08)	–0.078	(–0.19)	–0.138	(–0.37)
Any error correction	0.027	(0.65)	0.022	(0.58)	–0.374	(–0.87)	–0.428	(–0.97)
Lagged dependent variable	–0.118[c]	(–2.60)	–0.115[b]	(–2.45)	–0.429	(–1.22)	–0.314	(–0.92)
Extensive robustness tests	0.131[c]	(3.61)	0.139[c]	(3.99)	–0.170	(–0.53)	–0.118	(–0.36)
Inputs and outputs								
Binary measure of democracy	–0.032	(–0.81)	–0.013	(–0.32)	–0.261	(–0.72)	–0.018	(–0.06)
Attainability	0.029	(1.28)	0.039	(1.50)	0.259	(1.61)	0.511[b]	(2.23)
Subjectivity	0.166[c]	(3.23)	0.249[c]	(2.98)	1.754[c]	(3.38)	3.108[c]	(4.36)
Normatively ambiguous outcome	0.028	(0.59)	0.076	(1.09)	0.064	(0.18)	0.469	(0.92)
Outcome dummies	No		Yes		No		Yes	
Total analyses (n)	1,135		1,134		1,093		1,092	
Total studies (n)	583		583		574		574	
R^2	0.103		0.141		0.061		0.132	

[a]$p < 0.1$; [b]$p < 0.05$; [c]$p < 0.01$. Ordinary least squares regression with analysis (benchmark specification on a substantively unique outcome indicator) as the unit and errors clustered by study. The discipline dummy for Psychology is omitted from the table as there are only three analyses from that discipline.

Additional tests, shown in **Supplemental Appendix B**, offer variations on this protocol, including classical errors rather than clustered errors and the omission of normatively ambiguous outcomes. We also test additional specifications, including one with only the predictor of theoretical interest (a bivariate model). Models 1–2 are tested with the addition of the normatively adjusted *t*-value as a control in an attempt to further isolate threshold bias from prodemocracy bias. These robustness tests produce results that are very similar to those shown in **Table 4**.

The first set of predictors in **Table 4** classifies the discipline of the study (judged by the journal). We find that analyses published in political science, sociology, and interdisciplinary journals are more likely than analyses published in economics journals (the reference category) to show signs of threshold bias—though no more likely to show signs of prodemocracy bias. Perhaps economists

are less concerned with statistical significance than authors in other fields, as suggested by Brodeur et al. (2020).

The second type of predictor examines the impact of the journal in which a study was published. Biases are especially worrisome if they infect the field's most influential journals, whose work is most likely to be cited and to influence policy debates. We measure journal impact with CiteScore, reporting average citations per document for a journal over a 3-year period for all Scopus-listed journals. We impute the value 0 for 63 analyses in unlisted journals and apply a logarithmic transformation to the resulting variable (Ln[CiteScore+1]). Analyses in **Table 4** show that journal impact factor is negatively associated with threshold bias (but not prodemocracy bias). Perhaps these journals, which are often deemed by scholars to be of higher quality, place a lower premium on "significance stars" and have stricter policies vis-à-vis replication and other measures that might reduce p hacking.

A third predictor concerns time trends, as measured by year of publication. As it happens, more recent studies are less likely to report a statistically significant result and (though the relationship is very weak) a result suggesting that democracy is positively related to normatively desirable outcomes. Perhaps authors of newer studies are less worshipful of the golden calf of p-values, or perhaps recent attention to the importance of publishing null results in the social sciences (e.g., Nyhan 2015) has altered reviewer and editor behavior.

A fourth predictor concerns the breadth of a study, understood as the (log-transformed) number of outcomes contained in each study. (Each analysis receives a score equivalent to the total number of analyses contained in that study.) We find a negative relationship to both outcomes, which is somewhat troubling. It could be that when an author undertakes to test democracy's relationship to numerous outcomes, her objectives and incentives are different than when the outcome is singular. The greater the range, the greater the acceptability of mixed or largely null findings. It is not newsworthy if democracy's relationship to a single outcome falls short of expectations; but it is newsworthy if democracy's relationship to several outcomes falls short (e.g., Mulligan et al. 2004). Likewise, the greater the range, the greater the likelihood that an author will impose a single estimator, and perhaps a single specification, on all tests, leaving less wiggle-room for p hacking. Of course, such straitjacket designs may not be appropriate for all outcomes. Also, one cannot discount the possibility that authors who set out to test democracy's relationship across multiple outcomes are on a skeptical mission from the get-go and hence biased against the conventional view that democracy has good effects.

A fifth set of predictors focuses on research design. We measure whether a single cross-section is used, whether country or year fixed effects are employed, whether any error correction (anything other than classical errors) is applied, whether a lagged dependent variable is included as a covariate, and whether there are extensive robustness tests. If we were to find that weaker designs more often render statistically significant results, or results that suggest a conclusion that democracy enhances a normatively desirable governance outcome, we would have some cause for concern. One sign of this can be found in the results for a lagged dependent variable, which is negatively correlated with threshold bias and (more weakly) with prodemocracy bias. Along other parameters, the results of these tests do not seem problematic. Either there is no relationship with the outcome or the stronger research design is correlated with more significant results.

A sixth set of predictors relates to the amount of information that researchers draw on. Adding more countries or years to the sample should, ceteris paribus, reduce standard errors. Unsurprisingly, we find that a larger number of countries increases the likelihood of obtaining significant results and, also unsurprisingly (since democracy is positively related to outcomes in most analyses), higher normatively adjusted t-values. A longer time series is positively related to significant results but not normatively adjusted t-values.

A final set of predictors concerns the choice of input and output measures. Various democracy indices are employed to study democracy's relationship to governance, as shown in **Table 1**. One important decision is whether to code democracy in a binary or nonbinary fashion. Binary measures reduce all regime variability to a simple dichotomy, risking a loss of relevant information. However, tests in **Table 4** indicate that the choice of a binary or nonbinary indicator has no (or very little) impact on threshold bias or prodemocracy bias.

Many outcomes are represented in our sample ($n = 30$), and even more indicators ($n = 212$). Arguably, indicators that are more attainable (as explained above in the section titled A Database of Studies) should be more likely to show a relationship to democracy. Indeed, this is what we find. Attainability is weakly associated with passing the threshold of statistical significance and more clearly associated with rendering a normatively positive (prodemocracy) result. This is the expected finding.

Another measurement issue concerns indicators that are operationalized in a subjective fashion, whether by country experts, research assistants, or lay survey respondents. Examples include all recorded measures of corruption and most (63%) human rights measures. Here, it may be difficult to distinguish causal relationships from relationships that are, in part, a product of coding assumptions. For example, survey respondents' impressions of how prevalent corruption is within a country may be impacted by their impression of how democratic that country is, under the assumption that more autocratic countries must be more corrupt.

Indeed, we find that subjective measures are strongly related to both of our outcomes, suggesting both threshold and prodemocracy bias. Of course, it could also be that objective outcomes are easier for regime agents to manipulate than, say, expert opinions. If autocratic regimes manipulate statistics to enhance their scores on objective outcomes (Knutsen 2021), this would account for why subjectively coded outcomes show a stronger relationship to democracy.

A final measurement issue concerns our coding of outcomes as desirable or undesirable. Reasonable people can and do disagree about whether high government spending is good or bad for society, for example. As such, one might expect ambiguous outcomes to be less subject to prodemocracy bias. Yet, no obvious relationship appears in **Table 4**. This may be viewed as good news for those worried about prodemocracy bias, though it is certainly not conclusive.

Heterogeneous Effects Reconsidered

Table 3 displays results for 30 governance outcomes, ranked according to median t-values. By this measure, the outcomes at the top are more strongly enhanced by democracy. We have already noted several important caveats. Some outcomes are not well studied, some exhibit high variance, and some are of uncertain normative import. Findings for these outcomes should be regarded as provisional. What else can be inferred from our analysis of potential biases in a democratic direction? Are some outcomes more prone to bias?

Reassuringly, few predictors are strongly associated with prodemocratic findings in Models 3–4 of **Table 4**. If the discipline, journal, year of publication, breadth of study, and design are associated with more democratic effects, that relationship is slight and difficult to distinguish from chance.

However, subjectively coded outcomes are associated with a stronger democratic effect. Depending on how one views this factor, one may want to rethink the rank-ordering of outcomes in **Table 3**. Specifically, if one believes that subjectively coded outcomes are prone to error (in the direction of a larger democratic effect than is justified), then the optimistic t-values registered for outcomes like human rights and corruption should be reevaluated.

Precisely how much impact do various features associated with the discipline, journal, year of publication, breadth, design, and inputs and outputs of an analysis have on the heterogeneous effects registered in **Table 3**? To assess this question, we run a series of regression analyses that

replicate Model 4 in **Table 4**. Here, we focus on estimates for dummy variables measuring each outcome category (the 30 outcomes, minus one as a reference category). These analyses, displayed in **Supplemental Appendix B**, confirm that most outcome dummies are quite stable when the model is adjusted for various measurement and design characteristics. Exceptions include the human rights and corruption categories, whose coefficients are adjusted downward once we account for their use of subjective measures.

CONCLUSIONS

Our main objective in this study is to survey the published body of work on democracy's effects. While our ambit is not comprehensive—we omit measures of conflict and instability, for example—it is certainly more comprehensive than previous reviews of this literature.

We find that analyses of this topic published over the past two decades are generally optimistic. Relying on *t*-values (which provide a comparable measure of effect and uncertainty across a wide range of inputs and outputs), we find that democracy is usually associated with a normatively desirable outcome. In plain words, democratically governed countries appear to be better governed overall than autocratically governed countries.

We have also shown that this democratic dividend varies quite a bit across outcomes. In **Table 3**, 30 outcomes are ranked by their median *t*-values. (Median and mean *t*-values are highly correlated overall, but in some instances the latter are unduly influenced by one or two outliers.) It is instructive to see where democracy has had the largest positive reported effects, and where it has mattered less or perhaps even had a negative effect.

Finally, we wrestled with the problem of assessment. Is the literature on our topic to be believed? In what ways might it be biased?

Studying academic work on a subject by looking at published studies is a bit like studying icebergs by observing only the parts above water. One must make inferences about whole icebergs, even though the above-water portions are unlikely to be representative of the below-water portions. There is, first of all, a threshold bias (aka publication bias) favoring studies that can report statistically significant results. More serious is the potential problem of overall bias: Many researchers are probably trying to prove the same thesis, namely that democracy is good for governance.

To better understand the data-generating process, we ran a series of regression tests. We find indications that threshold bias is less common in studies published in economics journals, journals with a high Impact Factor, studies published more recently, and studies that explore a wide range of outcomes. Unsurprisingly, we also find that analyses with larger samples, with extensive robustness tests, and without a lagged dependent variable are more likely to clear the threshold. One result that gives us pause is the association between subjectively coded outcomes and statistically significant results. We take this up below, since it also pertains to our second outcome.

The second set of analyses, which focused on normatively adjusted *t*-values, provides insight into what sorts of studies are most likely to show that democracy has a positive effect on governance. Here, just a few strong patterns are discernible, which is encouraging. (It would be disconcerting if certain disciplines, journals, and types of studies were strongly associated with prodemocratic findings.)

One of these patterns concerns subjectively coded outcomes, which were also associated with beating the 95% threshold. Two explanations may account for this association, neither of which is very encouraging. First, it seems likely that subjectively coded outcomes are subject to a degree of circularity. For example, a country where corruption is thought to be pervasive is not likely to be regarded as very democratic, and vice versa. It is difficult to separate cause from effect, especially

if democracy is also coded in a subjective fashion (as it generally is). Another possible explanation is that subjectively coded outcomes are less susceptible to manipulation by autocratic leaders. China can manipulate its GDP statistics more easily than it can affect its score on Transparency International's Corruption Perceptions Index. Note that according to the first explanation one should trust subjectively coded outcomes less, while according to the second explanation one should trust them more. It is not clear which way the bias runs, but it is pretty clear that there is some sort of bias.

Another strong association exists between prodemocratic findings and the attainability of a policy objective. The apparent impact of democracy on governance is stronger with respect to outcomes that are fairly easy for governments to achieve, e.g., because they involve a change in a formal rule or a change in expenditures. Outcomes that are affected by many other things, like growth or infant mortality, are less likely to show a strong relationship to regime type (in the expected direction). This makes sense intuitively. Of course, it is somewhat deflating given that these distal outcomes are generally more important, substantively, than the proximal outcomes that governments can easily accomplish. We care a lot more about infant mortality than health expenditures, for example. Yet, it seems clear that democracy matters less for outcomes that matter more.

This does not mean that regime type is inconsequential. Bear in mind the immense range of policies and policy outcomes that a government has some influence over. In this study, we have surveyed 30 distinct outcomes captured by 212 indicators. Extant work on these topics suggests that democracy matters for most of these outcomes and that its impact is likely to be in a "positive" direction (toward policies and policy outcomes that most people would approve). This pattern fits with our sense that constitutional features of a polity affect a wide range of policy outcomes, albeit often in a marginal fashion. Because of its broad purview—not because of its impact on any particular outcome—democracy matters. That is why a wide-angle approach to democracy's effects is needed.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Achen CH, Bartels LM. 2017. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton Univ. Press

Brodeur A, Cook N, Heyes A. 2020. Methods matter: p-hacking and publication bias in causal analysis in economics. *Am. Econ. Rev.* 110(11):3634–60

Carbone G. 2009. The consequences of democratization. *J. Democracy* 20(2):123–37

Coccia M. 2020. How does institutional change of democratization affect the origin and diffusion of technological innovation across countries? *J. Econ. Social Thought* 7(2):60–91

Colagrossi M, Rossignoli D, Maggioni MA. 2020. Does democracy cause growth? A meta-analysis (of 2000 regressions). *Eur. J. Politcal Econ.* 61:1–44

Coppedge M, Edgell A, Knutsen CH, Lindberg SI, eds. 2022. *Why Democracies Develop and Decline*. Cambridge, UK: Cambridge Univ. Press

Doorenspleet R. 2019. *Rethinking the Value of Democracy: A Comparative Perspective*. London: Palgrave-Macmillan

Doucouliagos H, Ulubaşoğlu MA. 2008. Democracy and economic growth: a meta-analysis. *Am. J. Political Sci.* 52(1):61–83

Franck TM. 1992. The emerging right to democratic governance. *Am. J. Int. Law* 86(1): 46–91

Fukuyama F. 2014. *Political Order and Political Decay: From the Industrial Revolution to the Globalisation of Democracy*. London: Profiles Books

Gerber AS, Malhotra N. 2008. Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Q. J. Political Sci.* 3(3):313–26

Giavazzi F, Tabellini G. 2005. Economic and political liberalizations. *J. Monet. Econ.* 52(7):1297–330

Halperin M, Siegle J, Weinstein M. 2009. *The Democracy Advantage: How Democracies Promote Prosperity and Peace*. New York: Routledge

Knutsen CH. 2021. A business case for democracy: regime type, growth, and growth volatility. *Democratization* 28(8):1505–24

Li H. 2015. The Chinese model of development and its implications. *World J. Soc. Sci. Res.* 2(2):128–38

Lührmann A, Gastaldi L, Grahn S, Lindberg SI, Maxwell L, et al. 2019. *V-Dem annual democracy report 2019. Democracy facing global challenges*. V-Dem Inst., Univ. Gothenburg. **https://www.v-dem.net/media/filer_public/99/de/99dedd73-f8bc-484c-8b91-44ba601b6e6b/v-dem_democracy_report_2019.pdf**

McMann K, Seim B, Teorell J, Lindberg S. 2020. Why low levels of democracy promote corruption and high levels diminish it. *Political Res. Q.* 73(4):893–907

Mulligan C, Gil R, Sala-i-Martin X. 2004. Do democracies have different public policies than nondemocracies? *J. Econ. Perspect.* 18(1):51–74

Nyhan B. 2015. Increasing the credibility of political science research: a proposal for journal reforms. *PS Political Sci. Politics* 48(S1):78–83

Przeworski A, Limongi F. 1993. Political regimes and economic growth. *J. Econ. Perspect.* 7(3):51–69

Rodrik D. 1999. Democracies pay higher wages. *Q. J. Econ.* 114(3):707–38

Rothstein B. 2011. *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago: Univ. Chicago Press

Treisman D. 2020. Democracy by mistake: how the errors of autocrats trigger transitions to freer government. *Am. Political Sci. Rev.* 114(3):792–810