Letter

What Can We Learn about the Effects of Democracy Using Cross-National Data?

JONATHAN STAVNSKÆR DOUCETTE   *Aalborg University, Denmark*

More than 1,100 studies have been published that examine the effects of democracy using cross-national data since 2000. This article examines whether these analyses have sufficient statistical power to detect an effect of democracy. Using Monte Carlo simulation and examining consensus effects previously reported in the literature, the article finds that studies are only powered to detect very strong effects of democracy when examining countries over time. This raises questions about what sort of relationships can be analyzed using cross-national data.

INTRODUCTION

Since 2000, more than 1,100 studies have been published that examine the effects of democracy using cross-national data (see Gerring, Knutsen, and Berge 2022 for an overview). However, there have been no attempt to establish whether such analyses have sufficient statistical power to detect an effect of democracy. A lack of power can be problematic, as it implies a high probability of committing a false negative (Type II error). Even when estimates are statistically significant there is a risk of vastly overstating effect size (Type M error) and estimates may even have the wrong sign (Type S error) when studies are underpowered (Arel-Bundock et al. 2022; Gelman and Carlin 2014). This article seeks to shed light on this issue by using simulation to examine variation in the estimates for the effect of democracy. It finds that, with currently available data, analyses are likely only powered to detect strong and non-dynamic effects of democracy.

A staggering amount of factors have been theorized to be affected by democracy. However, this article is primarily focused on economic development for several reasons: first, it is the outcome that has been examined most frequently by the literature (Colagrossi, Rossignoli, and Maggioni 2020; Gerring, Knutsen, and Berge 2022, 367); second, there are good theoretical arguments for finding a substantial and positive impact of democracy (e.g., Baum and Lake 2003; Gerring et al. 2005; Knutsen 2012); third, data on GDP per capita are availability for more countries and for longer time spans than is the case with most other outcomes; and fourth, economic development vary more than most other outcomes studied by the literature, such as infant mortality or civil war.[1] Thus, if power is an issue for detecting a large effect of democracy on economic development, then it is likely to also present an issue for other outcomes.

Using the most extensive data available on democracy and GDP per capita and the standard two-way fixed effects (TWFE) estimator, I find that democracy must make countries around 16% richer or more for analysis to have sufficient power (80% power at $\alpha = 0.05$). This represents a large effect when compared to both prior estimates in the literature (e.g., Acemoglu et al. 2019; Colagrossi, Rossignoli, and Maggioni 2020; Knutsen and Wig 2015) and to the distribution of estimates from a multiverse analysis of the relationship between democracy and economic development.[2] I document a similar pattern for an alternative outcome, civil war, where democracy must decrease the risk of onset by 80% or more when compared to the average probability of civil war onset for analysis to be sufficiently powered. Moreover, if data are missing for a few countries, the true effect size must be very large to attain sufficient statistical power. For example, for datasets containing 75 countries, democracy must cause countries to be around 24% richer for the analysis to be well-powered. The consequences for power are not as arduous if data are missing for earlier time periods, such as prior to WWII, as long as the outcome changes slowly over time. For outcomes that vary significantly from year-to-year, missing early time periods also reduces statistical power substantially even when effects are large.

If an effect of democracy exist, it is likely to be dynamic and growing over time (see, e.g., Acemoglu et al. 2019). A common approach to modeling this are event-studies that include dummies for the relative time prior to and after democratization in addition to country and year fixed

Jonathan Stavnskær Doucette [ID], Associate Professor, Department of Politics and Society, Aalborg University, Denmark, jostdo@dps.aau.dk.

---

[1] Note that the results of this article does not depend on the actual relationship between democracy and economic development. The example is used as a baseline for a plausible (and large) effect of democracy with extensive data coverage.

[2] The multiverse analysis varies factors such as controls, democracy indicator, and the time period and region included in the data.

effects. I show that this further exacerbate power issues. Even when the average effect of democracy on economic development is large (corresponding to an average effect of around 16%), studies are only powered to detect long-run effects.[3] Using most of the new staggered difference-in-difference estimators that take issues with the TWFE estimator into account further increases power requirements (Chiu et al. 2023; Egerod and Hollenbach 2024). Thus, one should be cautious when interpreting analysis of dynamic effects. The lack of power also implies that it is not possible to detect actual deviations from parallel trends prior to treatment. In addition, my results indicate that analyses are, except in extreme cases, unlikely to be powered to detect interaction effects, as sufficient power is not reached even when the average effect is large (16%) and the difference between groups is huge (200%). Finally, I show that, conditional on researchers finding a significant result, estimates may be the wrong sign or several magnitudes too large if the true effect is small. This is problematic as the effects of democracy literature shows evidence of selection on significance (Gerring, Knutsen, and Berge 2022).

This article makes several contributions. First, its findings have implications for studies of the effects of democracy using cross-national data. Researchers within this field rarely, if ever, consider the power of the statistical tests they conduct. My results suggest that scholars relying on the standard TWFE estimator and cross-national regime data should think carefully about statistical power as it is likely to pose issues for their analysis unless the effect sizes they study are very large. This is true even if data are available for the whole population of countries across many years. This likely also applies to other institutional causes such as state capacity and party institutionalization (e.g., Andersen and Doucette 2022; Bizarro et al. 2018; Hegre, Bernhard, and Teorell 2020). This echoes recent studies that find generally low statistical power in economics and political science more broadly (Arel-Bundock et al. 2022; Askarov et al. 2024; Ioannidis, Stanley, and Doucouliagos 2017).

Second, I show how different design decisions and features of the data impact statistical power when using cross-national data. For example, studying interactions or using a sample that only includes countries from one continent are unlikely to yield credible estimates of the effects of democracy. Given that democracy is a staggered and dynamic treatment, scholarship that seek to address related issues with the TWFE estimator (e.g., de Chaisemartin and d'Haultfoeuille 2020; Goodman-Bacon 2021; Sun and Abraham 2021), should note that power becomes even more of an issue when using appropriate staggered difference-in-difference estimators. The lack of power is especially pertinent when testing for parallel trends prior to democratization, as tests are unlikely to pick up anything but very large divergences (see also Egerod and Hollenbach 2024; Roth 2022).

Third, I illustrate how scholars can use simulation to examine what minimum effect size is required for a

study of the effects of democracy to be informative (see also Black et al. 2022; Egerod and Hollenbach 2024 for inspiration in this regard). The inherent features of regime data, such as strong autocorrelation and clustered transitions, makes it hard to artificially create similar data. Simulating with real-world data might help researcher ascertain whether their proposed research design is actually powered to detect probable effects.

SIMULATION APPROACH

Statistical power depends on the level of statistical significance (usually set at $\alpha = 0.05$), sample size, effect size, number of units treated, and variability. Researchers studying the effects of democracy can only manipulate the number of countries and years in the sample in so far as there is additional data for some time periods or regions that can be collected. However, in the main analysis, I assume that the most extensive sample available is used. This covers the period 1800–2015 for around 180 countries (based on democracy data from Boix, Miller, and Rosato 2013[4] and logged GDP per capita data from Fariss et al. 2022a; 2022b). The variability of the dependent and independent variables and the number of units treated are also mostly outside the control of the researcher.[5] Given that one uses the most extensive sample available, this begs the questions: what kind of effect sizes can reliably be detected and how does this compare to the minimum effect size of interest?

To answer these questions, I conduct a simulation-based power analysis that vary the treatment effect in small increments using real-world panel data on democracy and economic development. The country-time-series for the outcome and treatment are separated and randomly combined into new artificial countries. The aim of the simulation approach is to find the minimum effect size of democracy on economic development that is reliably detectable using standard datasets and empirical approaches in the literature. Using actual panel data ensures that the features of the data match the features one would normally encounter when estimating cross-national regressions. Trying to simulate this kind of data is likely to significantly overstate the level of statistical power, as simulated data are unlikely to match the degree of autocorrelation, clustering, and non-randomness present in actual cross-national datasets (see Black et al. 2022; Egerod and Hollenbach 2024, 20–1). This is especially the case when studying regimes as individual countries rarely experience more than one or two transitions to or away from democracy. Thus, this approach assesses the uncertainty inherent in the designs used by the effects of democracy literature.[6]

---

[3] That is, effects that appear more than 10 years after a transition to democracy.

[4] As present in the V-Dem version 12 dataset (Coppedge et al. 2022).
[5] Or at least they can only be changed by choosing between different measures of democracy and GDP per capita that hopefully should capture the same phenomenon anyway.
[6] This is also relevant as data often contain (almost) the entire population of countries, thus making sampling-based uncertainty estimates less appropriate (Abadie et al. 2020).

Unfortunately, it is uncommon in the literature to report or discuss a minimum effect size of interest. Thus, I cannot readily compare the minimum effect size that can be detected with the kinds of effects scholars in the field would find theoretically and practically relevant. As an alternative, I do two things. First, I compare to reported effects in recent studies. Acemoglu et al. (2019) find that democracies are about 15% richer on average;[7] Colagrossi, Rossignoli, and Maggioni (2020) report a slightly lower difference of approximately 12%; Knutsen and Wig (2015) find GDP per capita grow about 0.42% faster per year compared to autocracies. However, relying on reported estimates risks overstating the actual relationship between democracy and economic development if studies are underpowered and there is selection on significance in the literature (Gelman 2019). Second, I compare with the effects found in a multiverse analysis that vary the factors which commonly differs between prior studies of democracy and economic development.[8] Democracies are about 9% richer on average, while the interquartile range of estimates goes from 5% to 13%. Democracies grow 0.4% faster annually, while the interquartile range of estimate goes from 0.27% to 0.53%.[9]

I assess the performance of statistical significance tests based on a panel of countries (i) observed in different years (t). I adopt standard power thresholds of 80% and 90% with a significance level of 0.05 ($\alpha = 0.05$). An often used approach in the literature is a linear regression of $Ln(GDP/cap)_{it}$ on $Democracy_{it}$ and country and year fixed effects (γ_i, δ_t). This is also termed the TWFE estimator. The tests are based on standard errors that cluster on countries. I summarize the specification as

$$Ln(GDP/cap)_{it} = \gamma_i + \delta_t + \beta\, Democracy_{it} + \epsilon_{it.}$$
(1)

I also run specifications that include a lagged outcome ($Ln(GDP/cap)_{it-1}$), in essence having growth as the dependent variable instead. I vary the baseline effect of democracy β in increments of 0.01 to find the minimum effect size that corresponds to a power level of 80% and 90%. The number of countries in the sample is 180 (C), which corresponds to the observed number of countries with data on both the democracy and the GDP per capita variable in at least 1 year. ϵ_{it} captures other time-variant factors that affect a country's economic development. I evaluate the variability of $\hat{\beta}$ as follows.

I simulate the steps outlined below ten thousand times and save the $\hat{\beta}$ from each repetition:

1. Construct a panel dataset of countries observed from 1800 to 2015.
2. Assign logged GDP per capita-year series to each country based on data from Fariss et al. (2022b).[10]
3. Randomly assign democracy-year series to each country based on data from Boix, Miller, and Rosato (2013).[11]
4. Multiply $Ln(GDP/cap)_{it}$ by an increasing β in each simulation run in years where $Democracy_{it}$ is equal to 1.
5. Estimate $Ln(GDP/cap)_{it} = \gamma_i + \delta_t + \beta Democracy_{it} + \epsilon_{it.}$
6. Save $\hat{\beta}$.

Steps 2 and 3 ensure that $Ln(GDP/cap)_{it}$ and $Democracy_{it}$ are uncorrelated in expectation. Thus, without step 4, estimates of β should center around 0 if the TWFE estimator is unbiased in this case. The average $\hat{\beta}$ for the simulations where β is set to 0 is –0.001, indicating that the procedure does remove any correlation between democracy and logged GDP per capita.

---

[7] Their estimated effect is dynamic and grows over time. According to their estimates, GDP per capita grows by around 1% per year after democratization, and around 20 years after the transition it remains around 20% higher. To get an average effect estimate, I first calculate the assumed effect size in each observed country-year observation where a country in the data is democratic ($years\text{–}democratic_{it} * 0.01$ if $years\text{–}democratic_{it} < 21$ and $democracy_{it} = 1$, and 0.2 if $years\text{–}democratic > 20$). Next, I average over the observed effect sizes and get an estimate of 0.15. This corresponds to a Cohen's D of 0.18 (or 0.47 when using the leftover variation in the outcome once country and year fixed effects are partialled out).

[8] Specifically, I vary the following things: (1) democracy indicator (Boix, Miller, and Rosato 2013; 2018; and the dichotomous versions of POLITY (Marshall and Jaggers 2013), Freedom House (FH 2023), and V-DEM (Lührmann, Tannenberg, and Lindberg 2018; Coppedge et al. 2022); (2) geographic region (I exclude the regions of the $e\_regiongeo$ variable from V-DEM [Coppedge et al. 2022] in turn); (3) time period included in sample (1786–, 1900–, 1950–, and 1970–); (4) controls (none, logged population size [from Coppedge et al. 2022, taken from Fariss et al. 2022b], and state capacity [the fiscal capacity measure from Coppedge et al. 2022]). Effects are estimated using the specification described in Equation 1 with and without a lagged dependent variable. The analysis returns 2,400 estimates of the effect of democracy on economic development.

[9] Analyses are based on the variables as present in the V-Dem version 12 dataset (Coppedge et al. 2022).

FINDINGS

I now evaluate how the power requirements of this approach vary as a function of (i) effect size, (ii) the size of the treatment group, (iii) the number of years in the dataset, (iv) the presence of dynamic effects, and (v) the presence of an interaction effect. These represent common differences between studies of the effect of democracy, as (i) some outcomes are more loosely connected to democracy (e.g., Leipziger 2024; Paglayan 2021), (ii) and (iii) occasionally outcome data are only available for some countries or periods (e.g., Stasavage 2005), (iv) the effect of democracy often materialize slowly over time (e.g., Acemoglu et al. 2019), and (v) many scholars are interested in how different factors interacts with the effect of democracy (e.g., Cox and Weingast 2018).

---

[10] The average logged GDP per capita in the sample is 1.33, and the standard deviation is 1.16.

[11] Democracy is present in 33% of country-years, and 68% of countries in the sample introduce democracy at some point.

Varying Effect Size

Figure 1 displays the results for the baseline specifications. Using the most extensive sample available, there is insufficient power to detect an effect of democracy that is as strong as the ones reported in Acemoglu et al. (2019), Colagrossi, Rossignoli, and Maggioni (2020), and Knutsen and Wig (2015). In fact, for studies to reach an 80% power level, the true effect of democracy must be above 0.15 or 0.005 when including a lagged dependent variable.

Consequently, studies are only powered to detect large differences between democracies and non-democracies. Given the natural limits on the number of units (countries) that can be included in cross-national analysis, scholarship is unlikely to have sufficient power unless one studies relationships where democracy has a strong effect and data have good coverage. In the Supplementary Material, I further show that this result is consistent across different choices available to researchers analyzing the effects of democracy. First, I find a similar pattern (see Figure A1 in the Supplementary Material) when using an interval-scaled measure of democracy using the *v2x_polyarchy* variable from V-Dem (Coppedge et al. 2022). In addition, I analyze an alternative outcome— civil war onset—that has received substantial attention

in the democratization literature (see the "Democracy and Civil War" section). Event variables usually have much less variation, and a score of 1 on these variables is often rare. As a result, this scenario reflects less ideal conditions for finding an effect of democracy. Yet, these are conditions that are common in the effects of democracy literature (Gerring, Knutsen, and Berge 2022). I find that the lack of power is severe in this case. Next, I examine how these results change in cases where researchers do not have data for all countries.

Varying the Number of Countries in the Sample

Figure 2 shows the relationship between the number of countries in the dataset, effect size, and statistical power. The lack of power quickly becomes more pronounced when data are missing for some countries. When data are only available for 125 countries, democracy must cause countries to be around 18% richer on average to reach an 80% power level. If data are only available for 75 countries, democracy must cause countries to be approximately 24% richer. As a result, lacking data for a number of countries further reduce the number of relationships one can study as effect sizes must be substantially larger for analyses to be powered.

FIGURE 1. Effect Size and Statistical Power
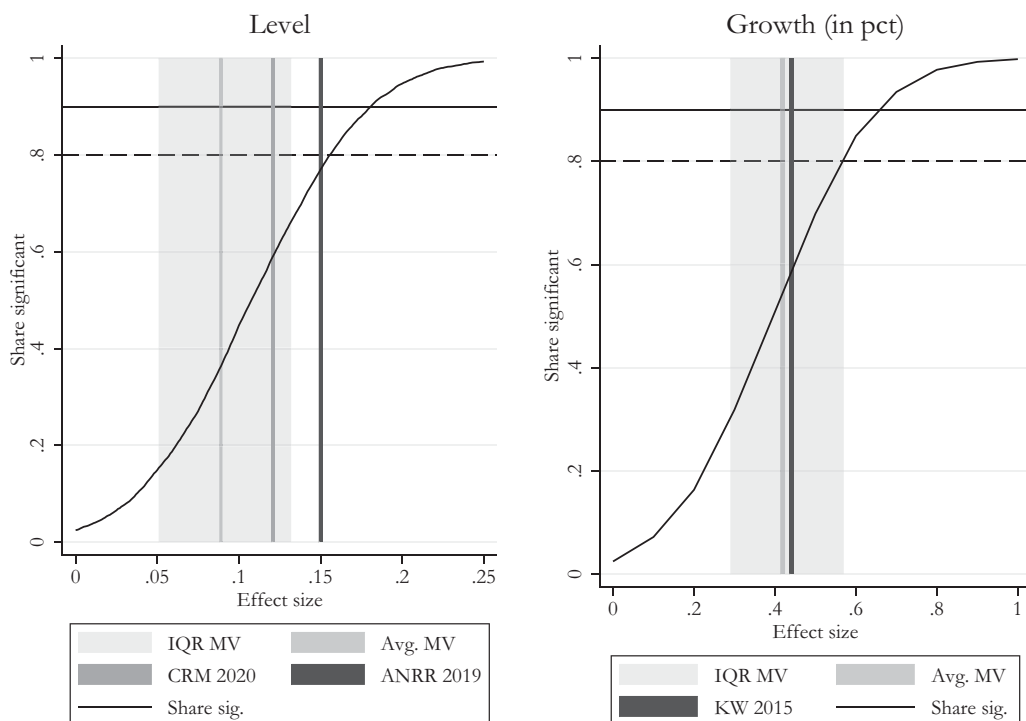
Note: Based on ten thousand repetitions per increment of effect size. The black line shows the share of estimates that are significant (in the right direction) at the 0.05 level across different effect sizes. The dashed vertical line corresponds to a power level of 80%, whereas the black line corresponds to a power level of 90%. The light gray area shows the interquartile range of multiverse estimates for the effect of democracy. The lightest gray horizontal bar shows the average effect across the multiverse estimates. In the left graph, the medium gray bar shows the estimate from Colagrossi, Rossignoli, and Maggioni (2020), whereas the black bar shows the estimate from Acemoglu et al. (2019). In the right graph, the vertical black bar shows the estimate from Knutsen and Wig (2015).

FIGURE 2. Countries in the Sample and Statistical Power
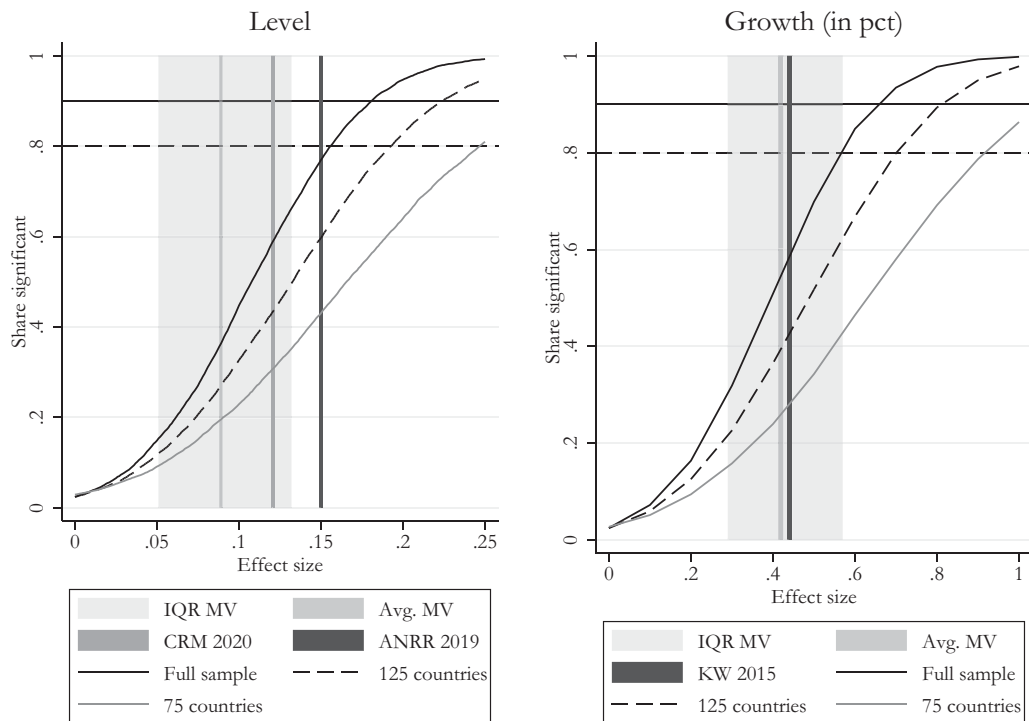
Note: Based on ten thousand repetitions per increment of effect size. The black line shows the share of estimates that are significant with the full sample, the dashed black line shows the share when there are 125 countries in the sample, whereas the gray line shows the share when there are 75 countries in the sample. The dashed vertical line corresponds to a power level of 80%, whereas the black line corresponds to a power level of 90%. The light gray area shows the interquartile range of multiverse estimates for the effect of democracy. The lightest gray horizontal bar shows the average effect across the multiverse estimates. In the left graph, the medium gray bar shows the estimate from Colagrossi, Rossignoli, and Maggioni (2020), whereas the black bar shows the estimate from Acemoglu et al. (2019). In the right graph, the vertical black bar shows the estimate from Knutsen and Wig (2015).
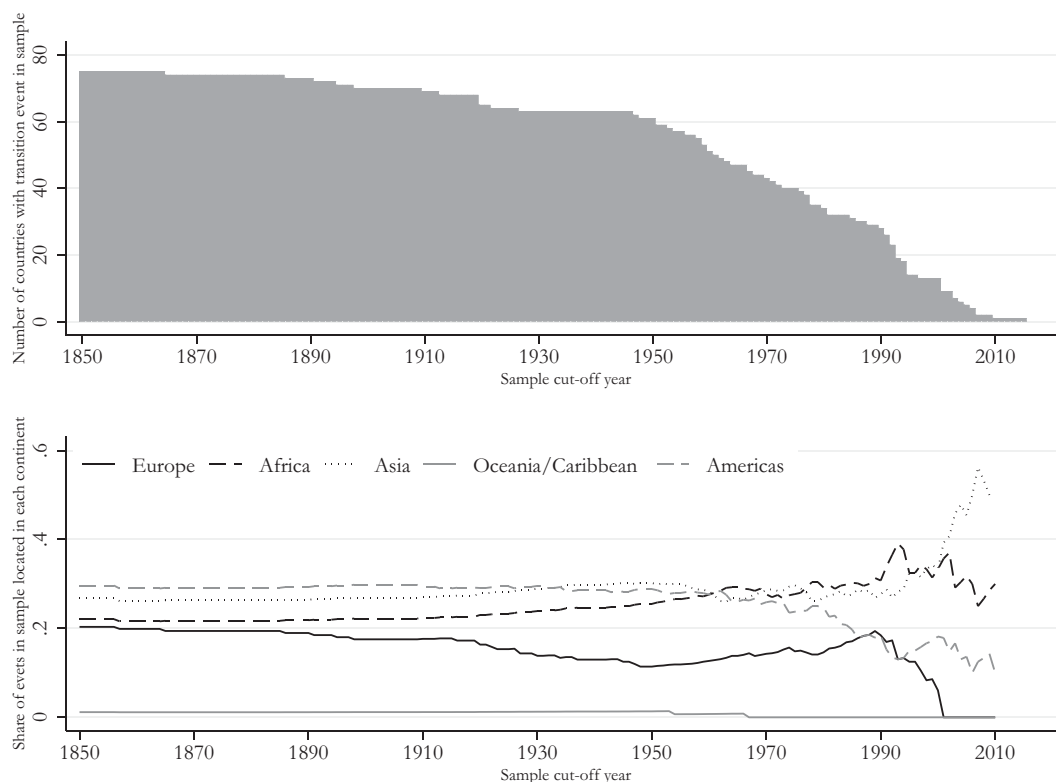
Varying the Number of Years Included

Missing data for a number of years often do not have equally dire consequences for statistical power when compared to missing countries, as standard errors are usually clustered on country and because missing countries can directly affect the size of the treatment group. However, as the upper graph in Figure 3 shows, a shorter time period can cause the size of the treatment group to shrink and reduce the number of countries to cluster on. Using the full sample, the treatment group includes 92 countries that transition to or away from democracy (i.e., changes treatment status). If data only include years after 1970, the treatment group almost shrinks to half as only 48 countries witness a regime change. This is because regimes tend to be sticky and change little year-to-year. In addition, as the lower graph shows, the effect one does find might reflect a very different geographic treatment group than the one found when using data for all countries.

Figure 4 displays the relationship between years included in the sample, effect size, and statistical power. As expected, the left graph, where growth is the outcome, shows that reducing the number of years included in the sample increases the effect size required for studies to be powered to detect an effect. However, somewhat surprisingly the relationship is actually reversed when the level of economic development is the outcome. Countervailing forces are at play here. On the one hand, reducing the number of years in the sample does decrease the number of treated units and the number of total units in the sample, which lowers power. On the other hand, when the level of development is the outcome and the time series for each country is short, the country fixed effects become very good at predicting the outcome (i.e., much less true when growth is the outcome), which lowers the standard error. Moreover, the reduction in the number of clusters (i.e., countries) is quite small. This increase in predictive power might offset the loss of (treated) units if the outcome only changes slowly over time. Nevertheless, in many cases, a shorter time period means less power, and it always implies that we are primarily studying the effects of democracy in specific parts of the world.

Dynamic Effects and Statistical Power

If an effect of democracy on economic development exist, it is likely to be dynamic and growing over time

1553

FIGURE 3. Consequences of Limiting the Number of Years for the Treatment Group

Note: The upper graph shows the number of countries that experience at least one switch in treatment status as a function of the number of years included in the sample. The lower graph shows the share of treatment events located in different parts of the world as a function of the number of years included in the sample. Europe and the Americas constitute a smaller share of the treatment group when the sample is limited to recent years, whereas Africa and Asia make up a larger share.

(see Acemoglu et al. 2019).[12] A common approach to modeling this are event-studies that include dummies for the relative years prior to and after democratization in addition to country and year fixed effects (often excluding a dummy for the year just before democracy is introduced). How does this alter the power requirements? To evaluate this, I use the estimates from Acemoglu et al. (2019), which indicate that GDP per capita grows after democratization in comparison with autocracies until about 20 years after democratization. At this point, democracies remain about 13%/16%/21% richer than autocracies (based on the multiverse analysis/Colagrossi, Rossignoli, and Maggioni 2020/and the minimum detectable effect [MDE]). To simulate this, I assume that β in the population grows by 0.0065/0.008/0.0105 each year after democratization and plateaus at 0.13/0.16/0.21 after 20 years. Figure 5 displays the distribution of estimates for the over time effect of democratization. It indicates that studies are generally underpowered to detect the dynamic effect of democracy. However, the power level does reach the standard threshold around 10 years after democratization if the true average effect of democracy is as large as the minimum detectable effect (16%). Thus, it may be

possible to recover large long-run effects. There are, however, two reasons to be cautious when studying dynamic effects. First, using most of the new staggered difference-in-difference estimators further increases power requirements (Chiu et al. 2023; Egerod and Hollenbach 2024). Second, if one lacks the power to detect short-term effects following democratization, it is also likely that one lacks the power to detect deviations from parallel trends prior to treatment (see also Roth 2022). Low power thus increases the risk that one misses the presence of nonparallel trends prior to democratization.

Interactions and Statistical Power

The effect of democracy may differ across groups, and as such, we might be interested in estimating this. Indeed, according to Colagrossi, Rossignoli, and Maggioni (2020), 38% of studies on democracy and economic development published since 2010 examine an interaction between democracy and another factor. However, as noted by, Gelman (2018), interactions increase the sample size required to detect an effect substantially.

To ascertain what implications interactions have for power when using cross-national data, I randomly

---

[12] It is probably likely that most effects of democracy are dynamic.

FIGURE 4. Years in the Sample and Statistical Power
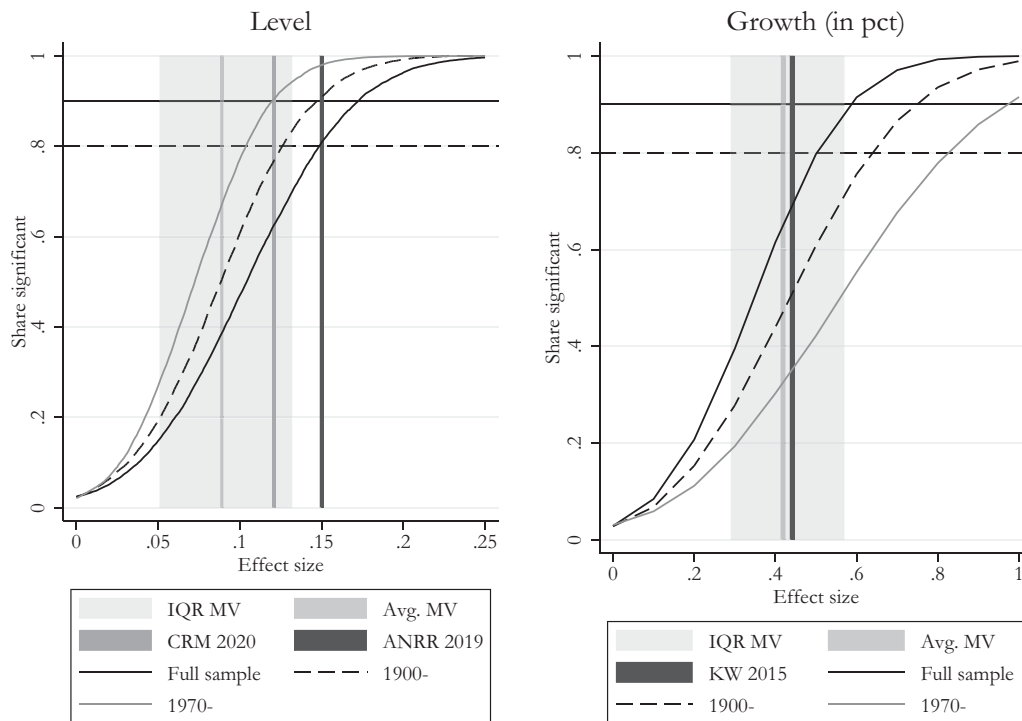
Note: Based on ten thousand repetitions per increment of effect size. The black line shows the share of estimates that are significant with the full sample, the dashed black line shows the share when only years after 1900 are in the sample, whereas the gray line shows the share when only years after 1970 are in the sample. The dashed vertical line corresponds to a power level of 80%, whereas the black line corresponds to a power level of 90%. The light gray area shows the interquartile range of multiverse estimates for the effect of democracy. The lightest gray horizontal bar shows the average effect across the multiverse estimates. In the left graph, the medium gray bar shows the estimate from Colagrossi, Rossignoli, and Maggioni (2020), whereas the black bar shows the estimate from Acemoglu et al. (2019). In the right graph, the vertical black bar shows the estimate from Knutsen and Wig (2015).

FIGURE 5. Dynamic Effects
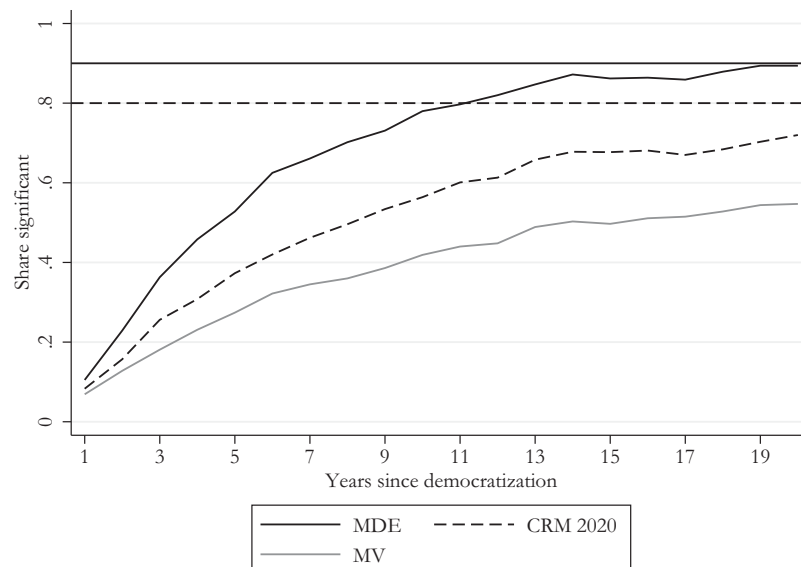
Note: Based on one thousand repetitions per increment of effect size. Dynamics are based on pattern reported in Acemoglu et al. (2019). The black line shows the share of estimates that are significant based on a dynamic effect corresponding to an average effect equal to the MDE (16%), the dashed black line shows the share based on the effect size reported in Colagrossi, Rossignoli, and Maggioni (2020), whereas the gray line shows the share based on the average effect size in the multiverse analysis. The dashed vertical line corresponds to a power level of 80%, whereas the black line corresponds to a power level of 90%.

FIGURE 6. Interactions and Statistical Power
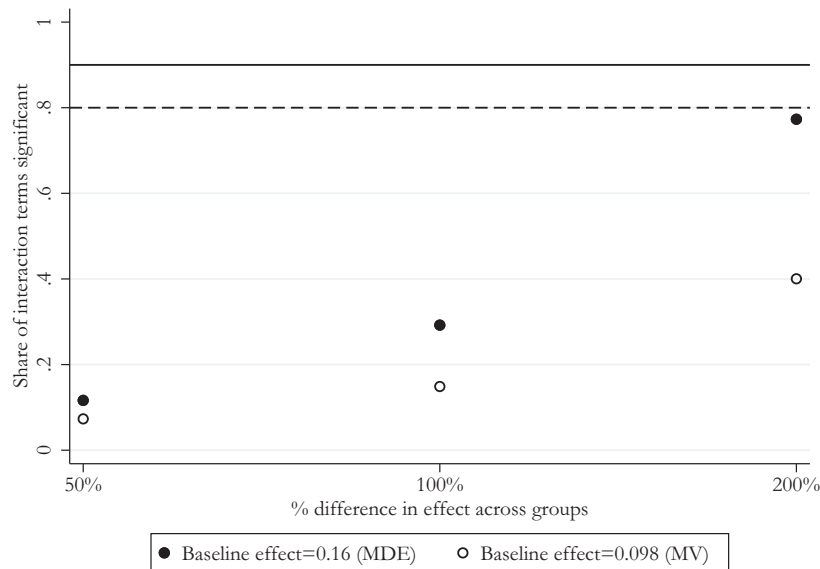
Note: Based on one thousand repetitions per combination of direct effect and interaction size. β is assumed to vary by Z_i. The average β is based on low scenario 0.098 (MV avg.) and the high scenario 0.16 (MDE). The TWFE models thus include an interaction term ($\beta_2(democracy_{it} \times Z_i)$) in addition to the term for democracy ($\beta_1 democracy_{it}$). Z_i is absorbed by country fixed effects.

assign countries into two groups and vary the size of the effect of democracy within each group according to three scenarios based on the size of the difference in effect size between groups.[13] To capture this, I include an interaction term between the democracy indicator and the group indicator in the baseline TWFE model. Figure 6 plots the share of interaction terms that are significant as a function of interaction effect size and baseline effect size. Even when the true baseline effect is very strong (MDE= β = 0.16) and the true interaction effect is very large (a 200% difference between groups), studies are not powered to detect an interaction effect. Thus, it is very unlikely that studies of democracy and economic development have sufficient power to study interactions.

Implications of Low Power

Besides increasing the risk of committing a false negative (Type II error), low power other and graver consequences for research. It increases the share of estimates with the wrong sign (Type S-Error) and it exaggerates the magnitude of the effect size (Type M-Error) (Gelman and Carlin 2014).

This is particularly problematic when power is low and there is selection on significance (Gelman 2019). Figure 7 illustrates this for studies of the effects of democracy. The left graph shows the share of

[13] (i) The high heterogeneity scenario with a null effect in one group and double the effect in the other group (a 200% difference), (ii) the medium heterogeneity scenario with a 100% difference in effect size between groups, and (iii) the low heterogeneity scenario with a 50% difference in effect size between groups.

significant results with the wrong sign (Type S-Error) at different true effects conditional on finding a significant result. When the true effect of democracy is relatively small (i.e., if democracies cause countries to be around 2.5% or less richer) and a significant effect is recovered, it is fairly likely that the result is in the opposite direction of the true effect. The right graph shows how much the effect is exaggerated, on average, compared to the true effect conditional on finding a statistically significant effect. When the true effect is small (0.05), significant results are exaggerated by 250% on average. This does drop when the true effect is larger, but even when the true effect is fairly strong (0.12), significant results are exaggerated by around 22% on average.

DISCUSSION

Taken together, these results suggest that analyses are only powered to detect strong effects of democracy. Thus, in a best case scenario, the absence of an effect of democracy for an outcome cannot be considered definitive proof that democracy has no effect on that outcome. Given the variability of the estimates and their sensitivity to the number of countries included and effect size, it is prudent to be cautious when interpreting significant effects as this might reflect noise given the low-powered nature of cross-national studies. In the worse case scenarios, statistically significant effects may be highly exaggerated or even in the wrong direction. Caution is further warranted as the effect of democracy is likely to be dynamic in many cases, which exacerbates power issues as appropriate estimators require additional statistical power.

FIGURE 7. Effect Size and Type S-Error and M-Error
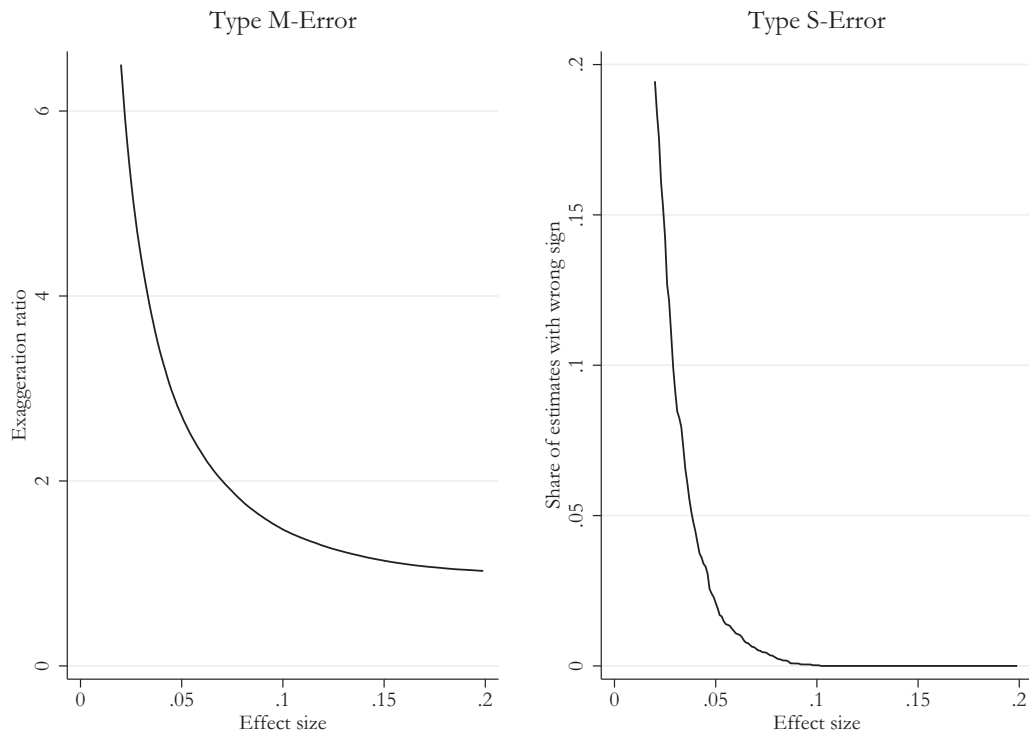
Note: Based on ten thousand repetitions per increment of effect size. Graphs are calculated based on simulations that find a significant result. The left graph shows how much the effect size is likely to be exaggerated, on average, at different true effect sizes (Type M-Error) (calculated as $\frac{\hat{\beta}}{\beta}$). The left graph shows the share of estimates that have the wrong sign at different true effect sizes (Type S-Error).

What can be done about this? First, a similar simulation exercise using real data and planned research designs should be undertaken before starting a study, which would reveal whether a planned analysis is likely to be informative or not. Here one could, for instance, consider whether additional statistical power can be gained by altering the design to include between country variation. However, researchers should be cautious and recognize that this likely trades bias from time-invariant confounders for power (this being a specific instance of the bias-variance trade-off). Second, one should be careful when interpreting the magnitude of effects found in cross-national analyses of the effects of democracy and recognize the uncertainty inherent in such estimates. In addition, one might supplement the analysis of the effects of democracy on an outcome by identifying additional implications of the theory that can also be tested. If the pattern is similar across outcomes, it raises confidence in the results. Moreover, in a small subset of cases there is data available on subnational variation in democratization (or at least on the theoretically relevant component of democracy) which can supplement the cross-national analysis (see, for instance, Grumbach 2023; Lankina and Getachew 2012 for data examples).

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit https://doi.org/10.1017/S0003055424001278.

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available in the American Political Science Review Dataverse at https://doi.org/10.7910/DVN/CLGI63.

CONFLICT OF INTEREST

The author declares no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The author affirms this research did not involve human participants.

Jonathan Stavnskær Doucette

REFERENCES

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2020. "Sampling-Based Uncertainty versus Design-Based Uncertainty in Regression Analysis." *Econometrica* 88 (1): 265–96.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James Robinson. 2019. "Democracy Does Cause Growth." *Journal of Political Economy* 127 (1): 47–100.

Andersen, David, and Jonathan Doucette. 2022. "State First? A Disaggregation and Empirical Interrogation." *British Journal of Political Science* 52 (1): 408–15.

Arel-Bundock, Vincent, Ryan Briggs, Hristos Doucouliagos, Marco Avina, and Tom Stanley. 2022. "Quantitative Political Science Research is Greatly Underpowered." I4R Dicussion Paper (6). Institute for Replication.

Askarov, Zohid, Anthony Doucouliagos, Tom D. Stanley, and Hristos Doucouliagos. 2024. "Selective and (Mis)leading Economic Journals: Meta-Research Evidence." *Journal of Economics Survey* 38 (5): 1567–92

Baum, Matthew, and David Lake. 2003. "The Political Economy of Growth: Democracy and Human Capital." *American Journal of Political Science* 47 (2): 333–47.

Bizarro, Fernando, John Gerring, Carl Henrik Knutsen, Allen Hicken, Michael Bernhard, Svend-Erik Skaaning, Michael Coppedge, et al. 2018. "Party Strength and Economic Growth." *World Politics* 70 (2): 275–320.

Black, Bernard, Alex Hollingsworth, Leticia Nunes, and Kosali Simon. 2022. "Simulated Power Analyses for Observational Studies: An Application to the Affordable Care Act Medicaid Expansion." *Journal of Public Economics* 213: 104713.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2013. "A Complete Dataset of Political Regimes, 1800–2007." *Comparative Political Studies* 46 (12): 1523–54.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2018. "Boix-Miller-Rosato Dichotomous Coding of Democracy, 1800–2015." Harvard Dataverse V3. https://doi.org/10.7910/DVN/FJLMKT.

Chiu, Albert, Xingchen Lan, Ziyi Liu, and Yiqing Xu. 2023. "What to Do (and Not to Do) with Causal Panel Analysis under Parallel Trends: Lessons from a Large Reanalysis Study." *SSRN Working Paper*.

Colagrossi, Marco, Domenico Rossignoli, and Mario Maggioni. 2020. "Does Democracy Cause Growth? A Meta-Analysis (of 2000 Regressions)." *European Journal of Political Economy* 61: 101824.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Nazifa Alizada, David Altman, et al. 2022. "V-Dem [Country-Year/Country-Date] Dataset v12." Varieties of Democracy (V-Dem) Project. https://doi.org/10.23696/vdemds22.

Cox, Gary W., and Barry Weingast. 2018. "Executive Constraints, Political Stability, and Economic Growth." *Comparative Political Studies* 51 (3): 279–303.

de Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogenous Treatment Effects." *American Economic Review* 110 (9): 2964–96.

Doucette, Jonathan Stavnskær. 2024. "Replication Data for: What Can We Learn about the Effects of Democracy Using Cross-National Data?" Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/CLGI63.

Egerod, Benjamin C. K., and Florian M. Hollenbach. 2024. "How Many is Enough? Sample Size in Staggered Difference-in-Difference Designs." OSF Preprint.

Fariss, Christopher, Therese Anders, Jonathan Markowitz, and Miriam Barnum. 2022a. "Latent Estimates of Historic Gross Domestic Product, GDP per Capita, Surplus Domestic Product, and Population Data Version 1." Harvard Dataverse V2. https://doi.org/10.7910/DVN/FALCGS.

Fariss, Christopher J., Therese Anders, Jonathan N. Markowitz, and Miriam Barnum. 2022b. "New Estimates of over 500 Years of Historic GDP and Population Data." *Journal of Conflict Resolution* 66 (3): 553–91.

FH. 2023. "Freedom House: Freedom in the World 2023." Report. New York, Freedom House. https://freedomhouse.org/report/freedom-world.

Gelman, Andrew. 2018. "You Need 16 Times the Sample Size to Estimate an Interaction than to Estimate a Main Effect." https://statmodeling.stat.columbia.edu/2018/03/15/need16/.

Gelman, Andrew. 2019. "Don't Calculate Post-Hoc Power Using Observed Estimate of Effect Size." *Annals of Surgery* 269 (1): 9–10.

Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9 (6): 641–51.

Gerring, John, Philip Bond, William Barndt, and Carola Moreno. 2005. "Democracy and Economic Growth: A Historical Perspective." *World Politics* 57 (3): 323–64.

Gerring, John, Carl Henrik Knutsen, and Jonas Berge. 2022. "Does Democracy Matter?" *Annual Review of Political Science* 25: 357–75.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.

Grumbach, Jacob. 2023. "Laboratories of Democratic Backsliding." *American Political Science Review* 117 (3): 967–84.

Hegre, Håvard, Michael Bernhard, and Jan Teorell. 2020. "Civil Society and the Democratic Peace." *Journal of Conflict Resolution* 64 (1): 32–62.

Ioannidis, John P. A., Tom D. Stanley, and Hristos Doucouliagos. 2017. "The Power of Bias in Economics Research." *The Economic Journal* 127 (605): F236–65.

Knutsen, Carl Henrik. 2012. "Democracy and Economic Growth: A Review of Arguments and Results." *International Area Studies Review* 15 (4): 393–415.

Knutsen, Carl Henrik, and Tore Wig. 2015. "Government Turnover and the Effects of Regime Type: How Requiring Alternation in Power Biases against the Estimated Economic Benefits of Democracy." *Comparative Political Studies* 48 (7): 882–914.

Lankina, Tomila, and Lullit Getachew. 2012. "Mission or Empire, Word or Sword? The Human Capital Legacy in Postcolonial Democratic Development." *American Journal of Political Science* 56 (2): 465–83.

Leipziger, Lasse E. 2024. "Does Democracy Reduce Ethnic Inequality?" *American Journal of Political Science* 68 (4): 1335–52.

Lührmann, Anna, Marcus Tannenberg, and Staffan Lindberg. 2018. "Regimes of the World (RoW): Opening New Avenues for the Comparative Study of Political Regimes." *Politics and Governance* 6 (1): 60–77.

Marshall, Monty G., and Keith Jaggers. 2013. "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2014." Center for Systemic Peace. https://www.systemicpeace.org/inscrdata.html.

Paglayan, Agustina. 2021. "The Non-Democratic Roots of Mass Education: Evidence from 200 Years." *American Political Science Review* 115 (1): 179–98.

Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Tends." *American Economic Review: Insights* 4 (3): 305–22.

Skaaning, Svend-Erik, John Gerring, and Henrikas Bartusevicius. 2015. "A Lexical Index of Democracy." *Comparative Political Studies* 48 (12): 1491–525.

Skaaning, Svend-Erik, John Gerring, and Henrikas Bartusevicius. 2018. "A Lexical Index of Electoral Democracy." Harvard Dataverse V6. https://doi.org/10.7910/DVN/29106.

Stasavage, David. 2005. "Democracy and Education Spending in Africa." *American Journal of Political Science* 49 (2): 343–58.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogenous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99.