

Appendices

A Data

A.1 Disruptive Capacity

My measure of non-elite capacity is the proportion of the working-age population employed in high-capacity industries. Compiling this was a tricky task. In all, the data come from five different sources: two pertaining to the denominator (the population aged 15 to 64), and three pertaining to the numerator (a count of the workers employed in manufacturing, mining, construction and transport).

The former involved a few challenges, since reported categories sometimes spanned the desired divisions.¹ For instance, out of the 1,014 observations in Mitchell’s historical dataset, 84 report at least one (or more) categories that straddle the distinction between working-age and not. These observations are reported at intervals, so the problem is greatly ramified by interpolation. To address this, I estimated the proportion of these categories that was working-age by examining the closest year in that country which did not report ambiguous categories, and used that proportion to estimate the working-age population in the otherwise troublesome year.

The latter task was even trickier, owing to considerable inconsistencies in the labour force categories reported across countries in Mitchell’s historical data (IHS) and the ILO’s contemporary data (ILO), and sometimes even within countries in different years.² The Groningen (GGDC) data is harmonized, but coverage is limited to 42 countries over 50 years. I considered the GGDC the most reliable source, but extended coverage in time and space by using the IHS and the ILO data.

The central challenges with these data were to (1) harmonize the classifications in the constituent datasets in order to combine these data usefully; (2) infer the number of high-capacity workers in years in which classifications straddled the desired divisions (e.g., where workers in mining and agriculture were reported together). This first task was too specific to detail here, but the crux of the problem was to decide how to deal with categories that combined subsectors of two larger categories. My approach was to classify these subcategories as an instance of the larger ambiguous category of which they were a part (e.g. if workers in utilities were reported alongside service sector workers, I classified these workers in the ambiguous category of ‘Manufacturing and Services’). The second task, then, was to decide how to handle ambiguous categories. One could drop these observations, but this is both inferentially and practically unsound. The reporting of ambiguous categories is probably associated with the size of the given categories, and these sizes figure centrally in my variable of interest. It also results in a very significant missingness problem in the early years, since observations cannot be interpolated across years that are missing. Instead, I estimated the proportion of these ambiguous categories that were high-capacity by looking at the nearest year within that country in which categories were reported separately. Table 5 shows this procedure in the example of the US case, where these early data come from IHS, but reporting is inconsistent. In 1850, 1860 and 1870 workers in commerce and transport are reported together. It is only in 1880 that they are reported apart. I infer the number of high-capacity workers in these

1. This applies exclusively to the Mitchell dataset available for early years; for later years annual and complete data are available through the UN.

2. The source code that generates my measure of the number of high-capacity workers is thousands of lines long.

early years from the proportion of commerce and transport workers that were in transport, in 1880. Totals in italics were computed from the original data.

Table 5: Commerce and Transport Workers in USA, from IHS

	commerce	transport	comm_trans	<i>HCAP.comm.trans</i>
1850	NA	NA	420	<i>174</i>
1860	NA	NA	780	<i>323</i>
1870	NA	NA	1350	<i>558</i>
1880	1220	860	<i>2080</i>	

In other words, to estimate for the number of high-capacity (transport) workers in the ambiguously reported years, I calculated the proportion of commerce and transport workers in 1880 ($\frac{860}{1220+860} = 0.413$), and then multiply this proportion by the number of workers reported in the ambiguous categories in 1850, 1860, and 1870 (e.g., in 1870, $1350 \times 0.413 \approx 558$).

A.2 GDP per capita

To create a composite GDP per capita series, I relied on two widely-used datasets: Angus Maddison’s Historical GDP series, and the Penn World Tables (Version 8.1). Though the Maddison dataset has coverage to the present era, merging it with the PWT is preferable for two reasons: to expand coverage to countries not in the Maddison dataset, and to incorporate the updated practices and data incorporated into Version 8.1 of the PWT.

The procedure I use to merge data across is fairly straightforward. First, I convert PWT data into 1990 US dollars, using instructions provided through correspondence with the Groningen Growth and Development Centre. I choose the output-based series since this seems to correspond more closely to the method used to calculate the Maddison data. I calculated a price level deflator for US dollars using *pl_gdpo* in order to compute what a 2005 US dollar was worth, in 1990 US dollar terms (about 0.72). In every country current-PPP GDP (*cgdp0*) in 1990 is expressed in 2005 US dollars, so I re-expressed this in 1990 US dollars using the deflator. I then used the *rgdpna* series to compute output-based growth rates for every year and country. I took the level calculated earlier, and used these growth rates to calculate the value of my desired series for years before and after 1990. This series gives an output-based GDP series in 1990 US dollars. On average, the resulting PWT series is correlated with the Maddison observations that overlap it at 0.8 within countries and over time, and at 0.9 across countries.

I decided that there was enough doubt in the appropriate correction to justify further adjustment as suggested in Miller (2012). That is to say, for countries in which I only had Maddison or PWT data, I relied on the available series. However, in the countries in which both were available, the Maddison series was typically available for years earlier than the first PWT observation. Simply appending the series to each other assumes away all measurement differences. If there *are* measurement differences, the resulting series will exhibit jumps at the point that data availability requires switching from one to the other. Instead, I extended the coverage of the PWT data by calculating an average factor of inflation for the Maddison data, using the three years of overlap closest to the first year

during which both series were available. In other words, I calculated the factor by which the Maddison data needed to be multiplied to equal the PWT data observed in those three years. I then used this factor of inflation to inflate all previously observed Maddison values. This adjusted the series in 108 countries. Reflecting the already close agreement between the two series, the average factor of inflation was 1.04 (the IQR of this inflation factor was 0.75 to 1.11) but at its most extreme the Maddison series was multiplied by 3.29 in one country and 0.16 in another.

A.3 Income Inequality

Income inequality features prominently in the three dominant models of the transition game, but measuring inequality accurately is a difficult task—one made substantially more difficult by the time period to which hypotheses about democratization pertain. I combined estimated Gini coefficients from three different datasets in order to maximize coverage. For the early years, I follow the precedent set by Ansell and Samuels (2014): I use Bourguignon and Morrison’s (hereafter BM) historical estimates of the income distribution to calculate Gini coefficients. These are obviously inexact, given that most countries are reported in groups, and estimates are reported at twenty-five year intervals.³ For more recent years, I relied on the *gini_net* series from the SWIID, and the *gini_best* from the SIDD. Where possible, missing values were interpolated.

My procedure for combining these estimates was similar to that described above. First, I extended the SWIID using data from the SIDD. In countries in which they overlapped *and* in which I had information from the SIDD about observations for which the SWIID was missing, I estimated a factor of inflation from the three overlapping observations closest to these missing observations. I used this to inflate the SIDD accordingly. This resulted in a gain of just 65 observations, or roughly 1% of the dataset.

Next, I used a similar approach to combine this composite series with BM’s historic estimates, using the three overlapping observations including the first year of overlap to estimate an average inflation factor. Multiplying the BM series by this average factor thus produced a long-run estimate of the income distribution. The final series has 7,589 observations and covers 182 countries.

This procedure means that this final series incorporates over-time information from BM and the SIDD, but takes its final level from the estimates given in the SWIID.⁴ In other words, the assumption is that over-time variation in the BM and SIDD captures important information about changes in the income distribution, even if the actual level is measured inaccurately. If this is an appropriate assumption, this procedure seems preferable to discarding the series and thus truncating the dataset.⁵

3. The original data are reported in nine deciles and two vintiles. I used this information to create an artificial income distribution from which I calculated a Gini coefficient.

4. Solt reports a distribution of estimates for each observation. This is intended to capture the uncertainty associated with measurement. I ignored this information in combining estimates across datasets, since there is no comparable information available for those countries and years in which the SWIID is not available.

5. Obviously, additional assumptions are involved. We assume that we can infer the magnitude of this difference in measurement from the three years of overlap between the two series, and that

this difference is a constant one. These are not likely to be true, but they are the best that can be done.

B Long-Run Multipliers

In any specifications which include a lag of the dependent variable on the right-hand side of the equation, independent variables will have both immediate (or short-run) effects on the dependent variable, as well as persistent (or long-run) effects. To see this, note that a single unit change in any of the independent variables at time t induces an initial adjustment in the dependent variable at time $t + 1$ of magnitude equal to its estimated coefficient (call it β). This is the short-run effect. In the absence of a lagged term, this short-run effect captures the entirety of an independent variable's effect on the level of the dependent variable. Where a lagged term is included among the estimators, however, the initial adjustment at time $t + 1$ will affect the level of the dependent variable at time $t + 2$. The magnitude of this effect is $\beta \times \alpha$, where α is the estimated coefficient associated with the lagged term. In turn, the new level of the dependent variable will have a knock-on effect at time $t + 3$, with magnitude $\beta \times \alpha^2$. This pattern persists, so the total effect of a single unit change at time t on the dependent variable at time $t + n$ is given by the geometric series $\beta + \beta\alpha + \beta\alpha^2 + \dots + \beta\alpha^{n-1}$. The long-run effect is derived by setting n to ∞ , in which case this series can be written as $\sum_{n=0}^{\infty} \beta\alpha^n$, which is equivalent to $\frac{\beta}{1-\alpha}$.

As De Boef and Keele (2008: pp.191-192) explain, calculating the standard error of this long-run estimate is not straightforward, since the long-run multiplier is the ratio of two (or more) coefficients.⁶ It is possible to directly estimate this uncertainty using the Bewley transformation. It is also possible to estimate it by simulation, which is how I proceeded in this paper. I simulated 5,000 draws from the appropriate variance-covariance matrix and calculated a distribution for the long-run multipliers of interest.⁷ I calculated a 95% confidence interval for the estimate by computing the 2.5th and 97.5th percentiles of this distribution. This is my preferred gauge of uncertainty, which is why I report confidence intervals rather than standard errors in the Tables. I approximated p-values for the estimate by examining whether the simulated 90%, 95% and 99% confidence intervals crossed 0.

6. When more than one lag of the dependent variable is included the long-run effect is equivalent to $\frac{\beta}{1-\alpha_1-\alpha_2-\dots-\alpha_n}$, where α_n is the coefficient associated with the n_{th} lag of the dependent variable. If more than one lag of the independent variable is included, the numerator changes to $\beta_1+\beta_2+\dots+\beta_m$, where β_m refers to the coefficient associated with the m_{th} lag of the independent variable.

7. I drew from the multivariate normal distribution defined by this matrix using the `mvrnorm()` function from the R package MASS.

C Predictions and Counterfactuals

C.1 Predictions

To generate predicted Polity2 (or Electoral Democracy) trajectories, I had to account for the inclusion of a lagged dependent variable in the estimation approach. Had I generated predictions using observed values (i.e., using observed Polity2 scores at time $t - 1$ to generate predicted Polity2 scores at time t), the cumulative effects of the other covariates would be essentially invisible. The overall trajectory would more or less indistinguishable from the observed trajectory.

Instead, to better understand what these models imply about the relationships between the independent variables and democracy, I proceed differently. I begin the predicted trajectory in year x , where x is the first year for which we have a Polity2 score and information for all of the independent variables in the model. This varies from country to country, depending on data availability. This yields a predicted Polity2 score for the year $x + 1$. Importantly, rather than using the observed Polity2 score in year $x + 1$ to predict the score in year $x + 2$, I use this predicted value. The new predicted score at $x + 1$ can in turn be used to predict the Polity2 score in year $x + 2$, $x + 3$, etc. In short, I predict values sequentially, rather than all at once. This approach is very similar to that used by Lin and Tomaskovic-Devey (2013) in their analysis of financialization and inequality.⁸

Of course, using previously predicted values to predict future values ramifies uncertainty. I do not describe this uncertainty in the paper, but in work not shown I made some attempt to calculate it by simulation. I estimated the errors associated with each prediction by drawing 100 samples from $N(\beta, \Omega)$, where β is the vector of coefficient estimates, and Ω is the associated variance-covariance matrix. However, I find that these results are not easy to interpret (i.e., the estimated errors are enormous), which I attribute to the high variance of estimates in fixed-effects models, and the fact that errors ramify with sequential prediction. I am working on a solution to these problems in other work. For now, I suppress discussion of predictive uncertainty in the paper, and I adopt the conventional strategy of relying on the underlying statistical significance of the estimates to know whether to trust the relevant counterfactuals.

C.2 Counterfactuals

In the paper I describe predicted values from two counterfactual scenarios applied to developing countries: (1) one in which GDP per capita increases as in the developed world, (2) one in which GDP per capita, disruptive capacity and landlord capacity evolve as in the developed world.

There might be several different and equally-reasonable ways to sketch these scenarios. My choice was to, first, plot a locally-weighted smooth to GDP per capita in all developed world countries, with time as the explanatory variable. I used this model to predict the average GDP per capita of a developed world country in any given year. This is the basis for counterfactual scenario (1). Specifically, to compute predicted trajectories for a given

8. Because Lin et al. (2013) model time-dependence using error-correction models rather than autoregressive distributed lag models, the technical details are slightly different.

developing country, I held all covariates at their observed values with the exception of GDP per capita, which took this counterfactual path. In other words, any given developing country would be assigned the ‘average’ per capita GDP of all developed countries in the relevant year, where ‘average’ means the value predicted by the locally-weighted smooth.

Scenario (2) derives from this first counterfactual, and employs the same method. I used the counterfactual trajectory of GDP per capita to generate an estimated path for the two capacity variables. Specifically, I fit locally-weighted smooths to two relationships: between disruptive capacity and GDP per capita (which is the explanatory variable), and between landlord capacity and GDP per capita. As above, I used these locally-weighted smooths to predict non-elite and landlord capacity as developed countries trawled the course of capitalist development. In this scenario, which I refer to in Figures 3 and 4, GDP per capita, non-elite capacity *and* landlord capacity follow a developed world pattern. Predicted trajectories are obtained as described above, using the counterfactual values for the relevant variables and with all other covariates held at their observed values.

D Model Fit

As is conventional, in all regression tables I report the adjusted- R^2 of the corresponding specifications. This is a crude measure of how well the model explains observed patterns, but in this case it is a particularly poor gauge of fit, for two reasons. First, in order to maximize coverage (and comparability to existing work), I estimated models on specification-specific samples. In the annual panel, for instance, the model which includes only a lag and a measure of GDP per capita is estimated on a sample more than twice the size of the sample used for the preferred model. This makes it impossible to compare the two statistics.⁹ Second, because all models follow the conventional strategy of including a lag of the dependent variable as well as year- and country- fixed-effects, there is not much residual variation by which different specifications can be distinguished. This is particularly true in the annual panel, given the high level of autocorrelation in the two dependent variables. Put another way, the inclusion of a lagged variable in all specifications means that the R^2 statistic captures how well the model in question explains a country's democracy score, given that country's score in the previous year. Unsurprisingly, most of the variation is mopped up by the antecedent observation. Thus, even when the first problem is corrected by estimating the different models on a consistent sample, the improvement in R^2 between the sparse and full specifications is negligible.

To better assess model fit, I considered a slightly distinct explanatory challenge. For each model, I generated a predicted trajectory for each country based only on the *initial* value of its dependent variable (plus the relevant covariates, and country- and year- fixed effects). Thus, though each model includes lags of the dependent variable as predictors, these lags were themselves predicted scores from previous years, and not Polity2 scores that were actually observed.

I compared these trajectories to the observed trajectory by computing two different statistics: the root mean square error and the mean absolute error. These two statistics are standard measures of error which capture how well the predicted series captures the year-to-year variation in the observed series. The RMSE squares deviations before averaging, so it weighs large mistakes more heavily than the MAE. Figure 12 displays how my preferred model compares to more parsimonious specifications.

These results affirm my choice of model. The preferred specification performs better than all sparser models. As compared to the empty model including only a lag and year- and country- fixed-effects, the preferred specifications results in something like a 8 to 10% improvement in fit. Note that the classic modernization model (adding only GDP per capita to the empty model) is not far better than the empty model.¹⁰

9. Notice that the first model is sparser, but the associated R^2 is higher in all panels. This is not surprising. The larger sample includes many more country-years from the 19th C., a period where the Polity2 index is significantly more invariant.

10. All these models have reasonably high absolute levels of fit. The empty model performs fairly well—it greatly reduces the error observed in models without fixed-effects at all. The large proportion of variation explained by these fixed-effects might be considered theoretically interesting. I prefer to understand it as a commentary on the difficulty of explaining the observed world. Either way, a fuller discussion of the predictive power of fixed-effects would have to consider the dangers of overfitting much more carefully than I have here (by, at minimum, estimating the models on a

different subsample of observations than that used for evaluating them). This is a fruitful avenue for future research to pursue, but it is beyond the scope of this paper.

E Robustness Tests

E.1 Alternative Models

Figure 6 shows that my main results are mostly robust to a variety of specifications. Each box in the grid illustrates the statistical significance of a particular variable (on the x-axis), in a particular model (on the y-axis), which are all grouped by type of robustness check (on the right-hand side).

The facet titled ‘Restricted Sample’ illustrates that the main results are the same when all models are estimated on a consistent sample, rather than on all observations available for a given specification (i.e., in the main text, sparser models are estimated on larger datasets). If they change, the estimated coefficients of disruptive capacity and landlord power are now significant at more exacting levels of statistical significance, and the substantive implications are the same.

Results grouped under ‘Cutoff’ show the results of running the preferred specifications at varying cut-off points for the minimum T_i (i.e., the minimum number of non-missing observations a country has to have to be included in the regression sample). My main results are not sensitive to changing the threshold that I use in the paper (i.e., the minimum number of complete observations that are present in a given country before I include it in the sample). The statistical significance of disruptive capacity fluctuates slightly, but only once drops below $\alpha = 0.05$.

Results grouped under ‘Specification’ show that other specifications do alter the DCAP result slightly: it is not statistically significant if no lag of the dependent variable is included, if the country fixed-effects are dropped. Neither of these decisions makes much sense. As I explain in the main paper, the obvious heterogeneity of countries in my sample and the perils of cross-country measurement recommend country fixed-effects. Similarly, and as I explain in a bit more detail in Section E.2 below, the strongly dynamic nature of democracy means that models without lagged dependent variables are inappropriate.

The DCAP result is significant (but sometimes inconsistently so) if random-effects are included, if additional lags of the dependent variable added, or if GDP is interacted with the Gini coefficient. The sensitivity of the result to the inclusion of additional lags in the already-truncated Electoral Democracy sample is likely just a consequence of its further truncation. All in all, none of these results give me strong reasons to doubt the main results.

Those grouped under ‘DCAP’ show that results are robust to alternative ways of measuring disruptive capacity. Two of these change the denominator to the labor force and the aggregate population, and the third changes the numerator to include public-sector workers (which greatly truncates the sample). My main results remain, except that the estimated impact of landlord power is unclear when public sector workers are included in the numerator (and Polity2 is the DV). Note that information on public sector employment is geographically and temporally restricted, so the sample is much smaller, which may very well explain this result. Finally, the results grouped under ‘Alternative DV’ show that the main results are robust to replacing V-Dem’s Electoral Democracy measure with its indices that capture other dimensions of democracy. Disruptive capacity is positively associated with these indices, landlord power negatively

so, and GDP per capita is not significantly associated (except that it is *negatively* associated with the egalitarian democracy index).

Figure 8 shows that my results are not driven by patterns in an outlying country (the y-axis shows which country is omitted). No matter which country I omit, results are substantively the same. GDP/capita is never statistically significant at conventional levels, though the estimate occasionally turns positive. The statistical significance of disruptive capacity fluctuates slightly in the models fit to V-Dem, but never drops below conventional levels (i.e., always significant at $\alpha = 0.05$).

E.2 To Lag or Not To Lag?

To furnish support for my view that it is preferable to include a lag of the dependent variable in my preferred specifications, I followed Keele and Kelly (2006). I ran Monte Carlo simulations within the parameter space that probably govern the data-generating process I observe. That is, I simulate a DGP that resembles (but simplifies) the data I gathered, where:

$$D_t = \alpha D_{t-1} + \beta DCAP_t + u_t \quad (1)$$

$$DCAP_t = \rho DCAP_{t-1} + e_{1t} \quad (2)$$

$$u_t = \phi u_{t-1} + e_{2t} \quad (3)$$

Here, D_t is the Polity2 score at time t , $DCAP_t$ is the disruptive capacity score at time t , and u_t are the model residuals. α and ρ govern the ‘stickiness’ of Polity2 and DCAP over time (i.e., the extent to which they are dynamic). I set these to the values I observe in the data, which suggest that both D and $DCAP$ are strongly dynamic processes (.958 and .984, respectively). β is thus the effect of disruptive capacity on democracy, which I vary in the interval around the β I estimate. ϕ indicates whether model residuals are autocorrelated, which I vary from 0 to 0.1 (i.e., no to mild autocorrelation). Following Keele et al. (2006), e_{1t} and e_{2t} are both drawn from a standard normal distribution. Unlike Keele et al. (2006), I consider multiple values of t . I set t to 20, 75, and 130, which spans the range of the panels in my sample (75 is their average length). Like them, I run 1000 repetitions at each permutation of ϕ , β and N .¹¹

I compared the performance of models with a lagged dependent variable (hereafter, LDVs) to Prais-Winsten regressions (which adjust the error structure to accommodate autocorrelation of the residuals, but do not explicitly model dynamics). I find, as the results in Keele et al. (2006) suggest, that in these kinds of data the LDV model outperforms both. Figure 9 plots the relative benefits of fitting an LDV model. I calculate this as the percentage improvement in the bias yielded by an LDV model relative to a Prais-Winsten model, where bias in either case is calculated as the absolute value of the percentage error in my estimate of β_{DCAP} . As the graph shows, throughout the parameter space, estimates from the LDV are more accurate. The median reduction in bias is about

11. One could extend this analysis to consider multiple exogenous variables and many units observed over time, but because I have no reason to believe this would change the conclusions, I did not do that here.

68.3% (the mean is 68.7%). The LDV performs better throughout the parameter space — at all levels of β , ϕ , and N .¹² This is good reason to trust results from the models I estimate over results from simple OLS.

E.3 Models of Transition and Consolidation

Figure 10 report results from to types of logistic regressions. Some scholars have argued that binary measures of a country’s democratic state are preferable to continuous measures. For one, the dynamics of partial and full democratizations (or their reverse) might differ, but linear regressions assume that the effect of a given covariate is constant at different points of the index. I do not prefer this approach. The theoretical model outlined in Section 3 suggests that the second of these concerns is unwarranted. Disruptive capacity should drive both partial and full democratizations. Put another way, non-elites will favor democratic concessions whether or not they lead to full democracy. More importantly, the use of a binary dependent variable introduces some significant inferential challenges. The use of country fixed-effects is ill-advised, since within-country variation on the dependent variable is rare. Most countries do not repeatedly transition to democracy from dictatorship or from dictatorship to democracy (Beck, Epstein, Jackman, and O’Halloran 2001). This is particularly the case when adopting the models recommended in the literature, which estimate separate models for transitions to and away from democracy (Przeworski 2000; Acemoglu, Johnson, Robinson, and Yared 2009; Miller 2012). This means that some amount of cross-sectional pooling is usually mandatory, which in turn carries some pernicious costs. The recent debate sparked by Robinson (2006), Acemoglu, Robinson, Johnson, and Yared (2008), and Acemoglu et al. (2009) was centrally concerned with the illegitimacy of making inferences from pooled estimators. The fact that more developed countries were more democratic, they argued, could not justify conclusions about the causal importance of development. At minimum, researchers were to restrict their attention to whether or not development improved a given country’s chances of democratizing.¹³ None of the responses to these two articles has disputed this point. Using a continuous index is thus also preferable because it eases the informational penalties of using past or future observations of the same country as a control. In sum, given the inferential difficulties associated with non-OLS estimation on data clustered in space and time, and the theoretical importance of modeling partial democratizations, I prefer the model described in Section 5.

Nonetheless, it is fair to wonder whether my results are robust to an alternative approach. I follow convention and estimate separate logistic regressions: the first set (labeled ‘Transition’) model the log odds of a transition at time t conditional on being a dictatorship at time $t - 1$, and the second set (labeled ‘Rollback’) model the log odds that a country will lapse into dictatorship at time t conditional on its being a democracy at

12. I do find that the Prais-Winsten estimates are more accurate than estimates from simple OLS. In other words, correcting the error structure is better than doing nothing at all. But the LDV outperforms both by a large margin.

13. Of course, it is important to clarify (as they do) that fixed effects do not permit causal inference. It is always plausible that a time-varying but unobserved third variable explains the association between two variables that vary together over time.

time $t - 1$. The specification is similar to the specification adopted for the main regressions, with a few key differences. First, models include regional dummies but do not contain the equivalent of country fixed-effects (i.e., they are not conditional logits). Democratic transitions are rare events, and the absence of significant within-country variation on the dependent variable imperils inference (Beck et al. 2001). As a further safeguard against country-level confounders, I include a dummy marking whether a country was ever a British colony, and a variable counting the previous number of democratic breakdowns it has suffered. Second, to account for time trends common to all regions I include decade dummies and a linear time trend. With year dummies, the models often fail to converge. Third, I measure the years that have elapsed since the country entered the risk pool with a cubic duration trend. These models are thus equivalent to event history models (Beck, Katz, and Tucker 1998), where the polynomial represents a Taylor Series approximation of the baseline hazard (Carter and Signorino 2010). These are thus the same models that Miller (2012) prefers in a recent paper on the modernization hypothesis.

The results from these models also suggest that class capacities matter. Both the ‘Transition’ and ‘Rollback’ models offer strong evidence that landlord capacity decreases the odds of seeing a transition under dictatorship, and increases the odds of seeing authoritarian rollback under democracy. In both cases, in the ‘Preferred’ models, landlord power is significantly associated with the probability of transition or rollback at $\alpha = 0.05$. Disruptive capacity matters, but in the models of rollback, it seems to matter in an unanticipated way: at low levels, an increase in disruptive capacity actually *increases* the chances of observing a coup in democracy (all else equal).¹⁴ At higher levels (when the number of high-capacity workers exceeds 20% of the working-age population), it starts to have the reverse and anticipated effect. In models of transition, there is no evidence of this kind of a squared effect — disruptive capacity is associated with the probability of seeing a transition to democracy in dictatorship at $\alpha = 0.10$. Note also that the level of GDP per capita seems a strong predictor of the likelihood of observing continued democracy under democracies. This is in keeping with earlier work which argued that development did not as much detonate democratization as safeguard it (Przeworski 2000). If this is accurate, it may invite some amendments to the model proposed at the outset of this paper, but I leave this for future research.

To ease interpretation, Figures 11 illustrate predicted probabilities from these four models, respectively. To make these quantities more meaningful, I have displayed the probability of observing the event in the subsequent ten years (rather than just the subsequent year). The cells correspond to various permutations of capacity. The first row displays values of disruptive capacity, and the first column displays values of landlord capacity (the percentile to which a value corresponds is noted in parentheses). Red tiles correspond to permutations that yield a high probability of the outcome, and blue tiles the opposite.

14. This is a suggestive result. It recalls the argument in O’Donnell (1979), which partly attributes the mid-twentieth century democratic reversals in Latin America to the *growth* of previously weak labour movements.

E.4 Unit Root Tests

Figure 13 shows that there is reasonable evidence that all series can be considered stationary. Evidence is not clear-cut for all series, but overall I consider it sufficient to justify the estimation approach I adopt in the paper. I ran a standard bevy of panel unit root tests on a balanced panel of varying sizes, employing the ‘plm’ package in R.¹⁵ In all variables and at all thresholds, a majority of unit root tests affirms the diagnosis of stationarity.¹⁶

15. To construct a balanced panel of a given number of countries N , I ranked all countries by the number of non-missing observations for a given series. I then chose the N th country in this list, identified the number of non-missing observations in it (say, X). To construct a balanced panel, I chose the X most recent observations from the countries ranked 1 to N on this list.

16. I chose the three panel unit root tests in the ‘plm’ package (Im, Pesaran, and Shin 2003; Levin, Lin, and James Chu 2002; Maddala and Wu 1999). I excluded a fourth (Hadri 2000), which proved unable to distinguish stationary from non-stationary series in the simulations that I ran. These simulations are not shown, but they are available on request.

F Interactions

One might wonder whether the model I elaborate in Section 3 is excessively parsimonious. Perhaps the effect of a given variable is actually conditional on the values of other variables, or maybe its relationship to democracy is non-monotonic? These hypotheses can be tested by adding interactions to the model I estimate, but this approach brings some dangers in train. Testing interactions liberally can encourage researchers to infer false meaning from chance relationships in the data — what Gelman and Loken (2013) calls ‘The Garden of Forking Paths’. Thus, a full consideration of these more complicated relationships merits more careful theoretical motivation than I can provide in this paper.¹⁷

This said, here I consider three possibilities. First, does too much disruptive capacity threaten democracy? That is to say: is the relationship between disruptive capacity and democracy non-monotonic? I argued that elites concede democracy when they face credible threats of disruption from non-elites, but it is possible that, if the threat from non-elites is too great, elites will actually redouble their commitments to dictatorship. This implies that the impact of disruptive capacity on democracy should be positive, at lower levels, but negative, at higher levels. As Figure 7 illustrates, I did not find any evidence that this was the case. The impact of disruptive capacity remains positive at all values of disruptive capacity (and, except at very low values, the result is always statistically noteworthy at conventional values). Nonetheless, further research should consider better ways of testing the postulate. I am limited to assuming that non-elites can threaten revolution when a very large number of them are clustered in mining, manufacturing, construction and transport. While this was the classic Marxist prophesy, none of our leading theories of revolution give us good reason to believe it.

Second, there are some historical reasons to suspect that the combination of a strong landed elite and strong non-elite bodes ill for democracy (e.g., Latin America in the 1960s and 1970s, or interwar Europe in the 1920s and 1930s). This expectation implies that landlords will react especially adversely to a mobilizing non-elite. Note that this is not the same as arguing that strong landed elites have a uniquely negative impact on democracy. This is what I already concluded in the section outlining my theoretical model. For the impact of disruptive capacity to change (or to even turn negative) when landlords are in power implies something more: that landed elites have less capacity for concessions than their industrial counterparts, perhaps, and thus turn with particular alacrity to authoritarianism when confronted with ascendant non-elites. Again, to test this, I interacted disruptive capacity with landlord strength. This revealed only limited evidence of a dynamic like this one. As Figure 7 illustrates, the impact of disruptive capacity is always positive at all levels of landlord power. It never turns negative. It is true that the estimate is not significant and is significantly closer to 0 at high levels of landlord power. This suggests that when landlords hold sway, a given change in the level of non-elite capacity is less likely to elicit democratic concessions. Of course, the imprecision of this estimate may simply reflect the fact that high disruptive capacity and high landlord power rarely co-occur, given how I have defined them. Better tests require a better formalization

17. Acemoglu and Robinson (2006) model this sort of approach. They derive their hypothesis that inequality should be non-monotonically related to democracy from a careful, formal consideration of what their assumptions imply.

of the expectation and certainly better measurement. This is surely a profitable area for future research.

Finally, some have argued that the disruptive powers of workers are elevated during and immediately after wartime, either because work stoppages will prove particularly threatening to the state, or because workers are especially willing to leverage latent disruptive capacities towards class conflict (Silver 2003: pp.140-141), or both. If this is correct, the positive impact of disruptive capacity on democracy should be highest during war. I find little evidence of this, however, when examining an interaction of disruptive capacity with a dummy indicating whether or not a country is locked in interstate conflict. If anything, the estimate is slightly *less* positive during war (and it is not significant in the Electoral Democracy models).¹⁸

18. Note that this may simply reflect the fact that, since war is rare, there is not sufficient evidence to establish how it matters.

Variable	Source	Coverage	Definition
Polity2 Score	PolityIV Project ¹⁹	1800-2014 ²⁰ , 201 countries ²¹	Annual democracy scores for all sovereign countries on a twenty-one point scale (i.e., the Polity2 series).
Electoral Democracy	V-Dem Version 6.2 ²²	1900-2015, 183 countries	Annual measure of the extent of electoral democracy in a given country. A continuous measure between 0 and 1.
Democracy (Binary)	Boix, Miller and Rosato ²³ ; Cheibub, Gandhi and Vreeland ²⁴ ; PolityIV Project	1800-2014, 230 countries	Annual democracy score for all countries on a binary scale. A country is coded as democratic when Boix and Cheibub agree. If this measure is missing, I treat a positive Polity2 score as indicating democracy.

19. <http://www.systemicpeace.org/inscrdata.html>

20. This refers to the range available in the original dataset, or , where applicable, the range available by virtue of combining information across datasets

21. ‘Countries’ here refers to countries that I count as distinct, which does not always exactly correspond to the distinctions maintained in the original data. For example, I code East and West Pakistan (1947-1971) as distinct from either West Pakistan (1972-present) and East Pakistan (1972-present), but not all datasets follow suit.

22. <https://www.v-dem.net/en/data/data-version-6-2/>

23. <https://sites.google.com/site/mkmtwo/data>

24. <https://sites.google.com/site/joseantoniocheibub/datasets/democracy-and-dictatorship-revisited>

Disruptive Capacity	Mitchell's International Historical Statistics, All Volumes ²⁵ ; International Labour Organization ²⁶ ; Groningen Growth and Development Centre ²⁷ ; United Nations World Population Prospects ²⁸	1754-2012, 177 countries	The number of workers in manufacturing, mining, transport and construction as a proportion of the population aged 15 to 64.
Landlord Capacity	Vanhanen's Democratization and Power Resources 2000 ²⁹ ; Vanhanen's Index of Power Resources 2007 ³⁰	1858-2007, 190 countries	The product of the percentage of land not in the hand of family farmers and the proportion of the population still in agriculture.
GDP per capita	The Maddison-Project ³¹ ; Penn World Tables 8.1 ³²	1 C.E.-2011, 192 countries	Gross Domestic Product per capita in 1990 1990 \$US, combined using algorithm to harmonize levels across the two sources.
Average Years of Education	Century of Education, Morrisson and Murtin ³³	1870-2010, 102 countries	Average years of education obtained by those aged 15 to 64.
Big City Population	Cross-National Time-Series Data Archive ³⁴	1815-2002, 221 countries	The per capita population living in cities of over 100,000 people.

25. <http://www.palgraveconnect.com/pc/archives/ihs.html>

26. <http://www.ilo.org/ilostat>

27. <http://www.rug.nl/research/ggdc/>

28. <http://esa.un.org/unpd/wpp/DVD/>

29. https://services.fsd.uta.fi/catalogue/FSD1216?study_language=en

30. <https://services.fsd.uta.fi/catalogue/FSD2420?lang=en>

31. <http://www.ggdc.net/maddison/maddison-project/home.htm>

32. <http://www.rug.nl/research/ggdc/data/pwt/>

33. By request

34. <http://www.cntsdata.com/>

Income Inequality	Bourguignon and Morrison ³⁵ ; SIID, Babones and Alvarez-Rivadulla ³⁶ ; SWIID, Solt ³⁷	1820-2013, 182 countries	Income inequality as measured by the Gini coefficient (net of government taxes and transfers, where distinguishable).
Strike Frequency and Strike Volume	Mitchell's International Historical Statistics, All Volumes ³⁸ ; International Labour Organization ³⁹	1881-2008, 138 countries	Strike frequency is the number of strikes (or lock-outs) scaled by some measure of the size of the eligible population. Strike volume is the number of days lost to strikes scaled by the same measure. My preference is to use a measure of the working-age population.
Union Membership	Visser's Historical Data ⁴⁰ ; International Labour Organization ⁴¹ ; Visser's ICTWSS ⁴²	1885-2012, 59 countries	The number of active union members scaled by some measure of the size of the eligible population. My preference is to use a measure of the working-age population. Data are combined across sources by using the same algorithm used to harmonize the GDP per capita data. I consider the ICTWSS to be the most reliable source.

35. By request

36. By request

37. <http://myweb.uiowa.edu/fsolt/swiid/swiid.html>

38. <http://www.palgraveconnect.com/pc/archives/ihs.html>

39. <http://www.ilo.org/ilostat>

40. By request

41. <http://www.ilo.org/ilostat>

42. <http://www.uva-aiaa.net/207>

Table 7: Short-Run Estimates, Polity2

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Lagged Dep. Var</i>						
Polity2 Score _{<i>t</i>-1}	0.911** (0.010)	0.916** (0.007)	0.910** (0.010)	0.916** (0.008)	0.928** (0.007)	0.899** (0.014)
<i>Short-Run Impact</i>						
Disruptive Capacity _{<i>t</i>-1}	0.263** (0.090)		0.234** (0.090)			0.421** (0.149)
Landlord Power _{<i>t</i>-1}		-0.411** (0.110)	-0.495** (0.140)			-0.587** (0.173)
GDP per capita (Log) _{<i>t</i>-1}				0.207 (0.237)		-0.127 (0.391)
Growth Rate _{<i>t</i>-1}				-0.000 (0.110)		0.010 (0.188)
Educational Attainment _{<i>t</i>-1}				-0.264 (0.327)		0.869 (0.555)
Urbanity _{<i>t</i>-1}				0.263* (0.121)		-0.231 (0.154)
Income Inequality _{<i>t</i>-1}					0.037 (0.089)	0.212 (0.155)
Polity2 Score (Reg. Avg) _{<i>t</i>-1}						0.556* (0.231)

This table reports the short-run estimates from which the results in Table 3 are derived. For an explanation of how one derives long-run estimates from short-run estimates, and what long-run estimates mean, see Section B of the Appendix.

Table 8: Short-Run Estimates, Electoral Democracy

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Lagged Dep. Var</i>						
Electoral Democracy _{<i>t</i>-1}	0.938** (0.007)	0.938** (0.006)	0.940** (0.008)	0.940** (0.007)	0.936** (0.006)	0.925** (0.009)
<i>Short-Run Impact</i>						
Disruptive Capacity _{<i>t</i>-1}	0.089+ (0.049)		0.089+ (0.046)			0.210* (0.082)
Landlord Power _{<i>t</i>-1}		-0.186** (0.061)	-0.200* (0.080)			-0.451** (0.113)
GDP per capita (Log) _{<i>t</i>-1}				0.200 (0.152)		-0.046 (0.249)
Growth Rate _{<i>t</i>-1}				0.118 (0.094)		0.178 (0.185)
Educational Attainment _{<i>t</i>-1}				0.152 (0.231)		0.899* (0.436)
Urbanity _{<i>t</i>-1}				0.082 (0.072)		-0.211+ (0.108)
Income Inequality _{<i>t</i>-1}					0.094+ (0.055)	0.161 (0.103)
Electoral Democracy (Reg. Avg) _{<i>t</i>-1}						0.541* (0.221)

This table reports the short-run estimates from which the results in Table 4 are derived.

Table 9: Long-Run Estimates, Polity2 At Varying Interval

	(1)	(2)	(3)	(4)	(5)
<i>Long-Run Multiplier</i>					
Disruptive Capacity	4.168** [1.35,6.94]	3.992** [1.56,6.27]	3.593** [1.08,6.25]	4.539** [1.58,7.69]	3.042 [-1.4,7.37]
Landlord Power	-5.774** [-10.07,-2.44]	-5.927** [-9.92,-2.67]	-6.612** [-10.84,-2.67]	-5.571* [-10.01,-0.77]	-7.662** [-12.44,-3.08]
GDP per capita (Log)	-1.219 [-8.77,6.76]	0.846 [-6.21,8.34]	0.441 [-6.41,7.5]	6.731 [-3.98,16.58]	10.886 [-3.25,24.11]
Growth Rate	0.094 [-4.02,3.72]	-1.611 [-5.11,1.63]	-1.468 [-4.75,1.62]	1.041 [-2.25,4.49]	0.507 [-3.52,4.48]
Educational Attainment	8.681 [-2.16,19.67]	4.862 [-5.49,15.13]	9.304+ [-1,19.62]	1.753 [-11.54,14.78]	0.291 [-15.14,15.27]
Urbanity	-2.287 [-5.23,0.81]	-1.978 [-4.76,0.97]	-2.685+ [-5.65,0.22]	-1.781 [-6.23,2.97]	-1.522 [-6.67,3.96]
Income Inequality	2.061 [-0.99,5.16]	2.097 [-0.84,4.75]	2.536+ [-0.28,5.33]	-0.005 [-3.44,3.41]	1.147 [-2.54,4.88]
Polity2 Score (Reg. Avg)	5.471** [1.18,9.69]	3.899+ [-0.17,7.3]	0.968 [-3.58,5.13]	1.302 [-3.97,6.18]	-0.920 [-5.77,3.29]
<i>Model Info</i>					
Observations	4,437	913	487	319	187
Countries	64	66	72	78	67
Range	1871-2003	1877-2007	1882-2012	1892-2012	1902-2002
Avg. N_i	69.3	13.8	6.8	4.1	2.8
Country-Level	FE	FE	FE	FE	FE
Year-Level	FE	FE	FE	FE	FE
Adj. R^2	0.869	0.486	0.293	0.109	-0.025

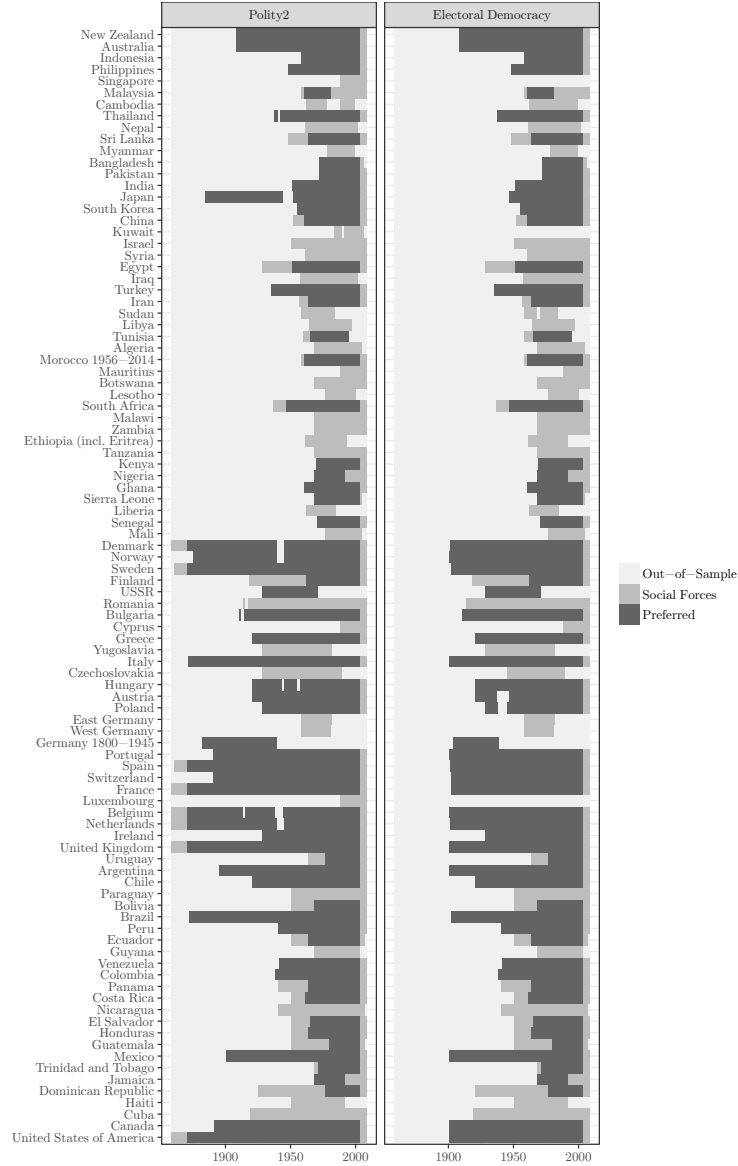
This table reports long-run estimates when the preferred model is estimated on a dataset of varying intervals between observations, and when the dependent variable is the Polity2 score (Figure 2 shows some of this information graphically). Column 1 is the annual sample, Column 2 is the 5-year, Column 3 is the 10-year, Column 4 is the 15-year and Column 5 is the 20-year sample. In the main, my results of interest survive. Disruptive capacity is a positive and statistically significant predictor of Polity2 at all but a twenty-year remove, and landlord power is a positive and statistically significant predictor of Polity2 at all removes. Note that the sample size diminishes dramatically as the interval between observations increases. (Indeed, the twenty-year model fits the data very poorly. Results are based on within-country variation, but the average country has only a couple of observations.)

Table 10: Long-Run Estimates, Electoral Democracy at Varying Interval

	(1)	(2)	(3)	(4)	(5)
<i>Long-Run Multiplier</i>					
Disruptive Capacity	2.812** [0.69,5.08]	1.877* [0.07,3.58]	1.799+ [-0.02,3.7]	2.296* [0.06,4.69]	1.538 [-0.77,3.73]
Landlord Power	-6.022** [-8.82,-3.31]	-5.674** [-8.52,-3]	-5.256** [-7.74,-2.69]	-4.533** [-8.02,-0.71]	-5.534** [-9.03,-2.03]
GDP per capita (Log)	-0.675 [-7.52,5.7]	-0.113 [-5.68,5.23]	-0.086 [-5.05,4.4]	3.482 [-3.7,9.8]	5.971 [-2.08,13.31]
Growth Rate	2.361 [-2.52,7.24]	-1.166 [-3.94,1.3]	-0.933 [-3.03,1.03]	0.154 [-2.32,2.64]	1.773 [-0.83,4.24]
Educational Attainment	11.938* [0.67,25.13]	5.001 [-5.17,15.3]	8.381+ [-0.19,16.83]	-0.661 [-13.78,12.09]	-0.839 [-13.76,11.92]
Urbanity	-2.809+ [-5.72,0]	-2.175 [-4.81,0.66]	-2.107+ [-4.3,0.01]	-1.456 [-5.04,2.17]	-1.744 [-5.57,2.14]
Income Inequality	2.160 [-0.49,5.11]	2.265* [0.16,4.48]	2.142* [0.03,4.21]	1.316 [-0.91,3.51]	2.413+ [-0.37,5.14]
Electoral Democracy (Reg. Avg)	7.274** [1.65,12.46]	4.628+ [-0.21,8.82]	2.432 [-2,6.03]	1.784 [-2.82,5.69]	1.297 [-3.32,5.3]
<i>Model Info</i>					
Observations	4,013	830	447	289	176
Countries	64	66	72	78	67
Range	1901-2003	1907-2007	1912-2012	1922-2012	1922-2002
Avg. N_i	62.7	12.6	6.2	3.7	2.6
Country-Level	FE	FE	FE	FE	FE
Year-Level	FE	FE	FE	FE	FE
Adj. R^2	0.926	0.623	0.475	0.268	0.250

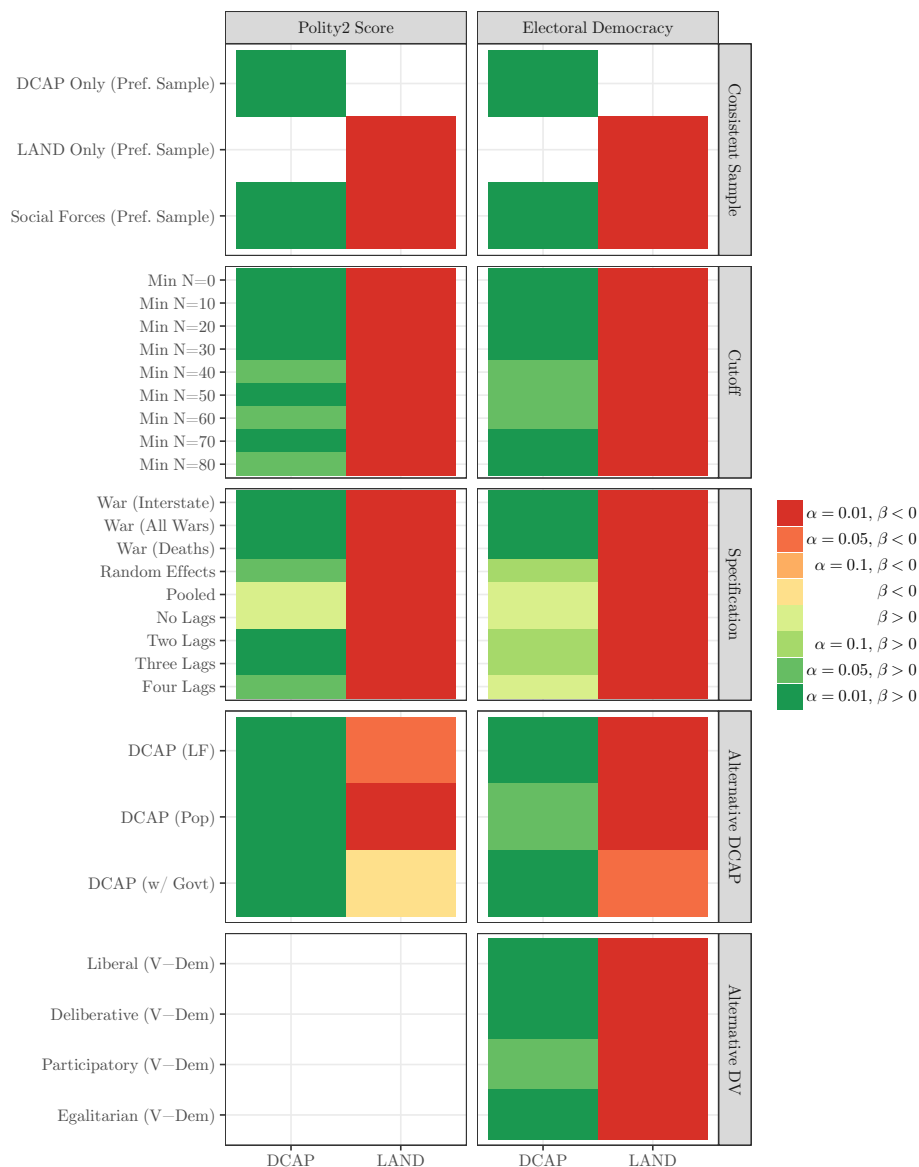
This table reports long-run estimates when the preferred model is estimated on a dataset of varying intervals between observations, and when the dependent variable is the V-Dem measure of electoral democracy (Figure 2 shows some of this information graphically). Again, the results are mostly consistent with my expectations. Disruptive capacity is a positive and statistically significant predictor of Electoral Democracy at all but a twenty-year remove (albeit at $\alpha = 0.10$ in the 10-year sample). Landlord power is a negative and statistically significant predictor of Electoral Democracy in all samples.

Figure 5: Country-Years in Sample



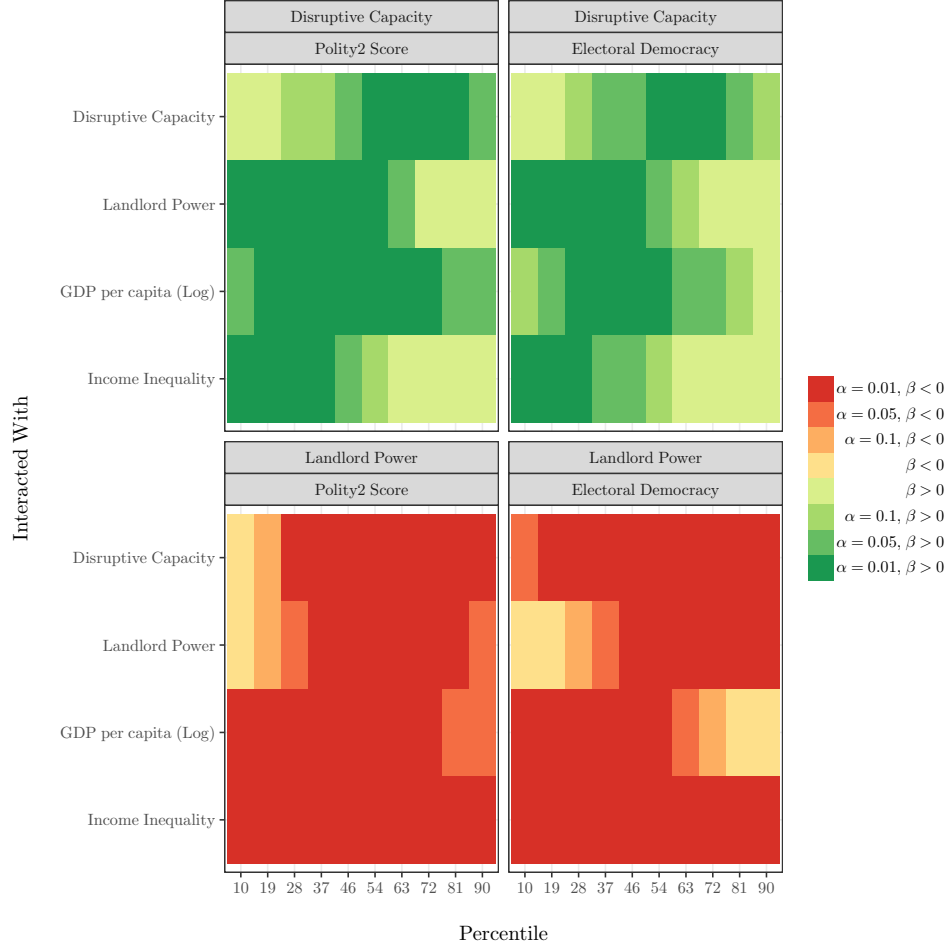
This graph plots the countries and years that are in the sample for my ‘social forces’ model (Column 3 of Tables 3 and 4), and in my preferred sample (Column 6). The data are concentrated in the second half of the 20th century. Only in advanced countries and in some Latin American countries are there substantial data available earlier. Moreover, since V-Dem starts in 1900, it is only in the Polity2 models that my sample spans the late 19th century.

Figure 6: Robustness Checks



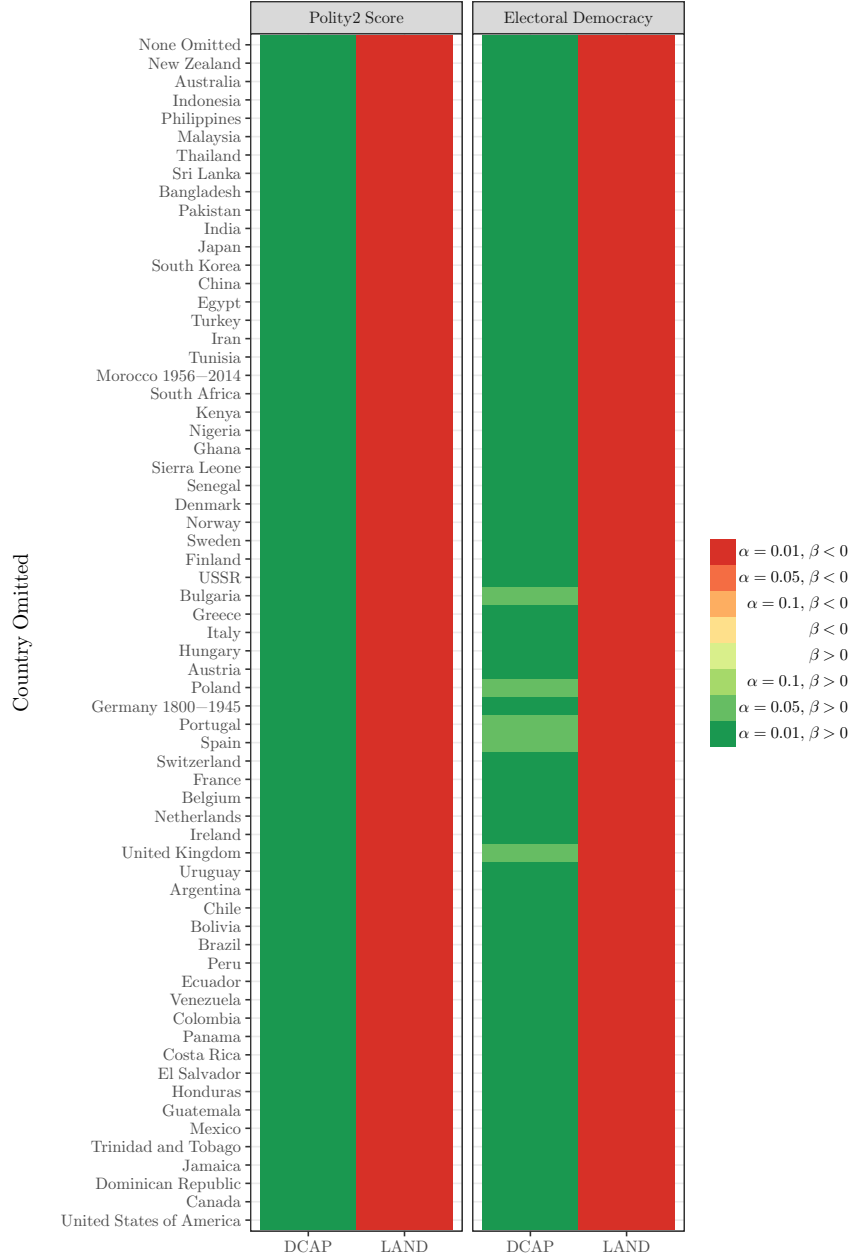
This graph reports the statistical significance of the estimated impact of disruptive capacity and landlord power, across a wide array of robustness checks on the annual sample. I have grouped these by type. The main takeaway is that my results are robust to a variety of alternative specifications. Where they are no longer statistically significant (e.g., the estimated impact of disruptive capacity in the 'Pooled' specification), there is good reason to favor my preferred model. For a full discussion of these results, see Section E in the Appendix.

Figure 7: Estimates from Models with Key Interactions



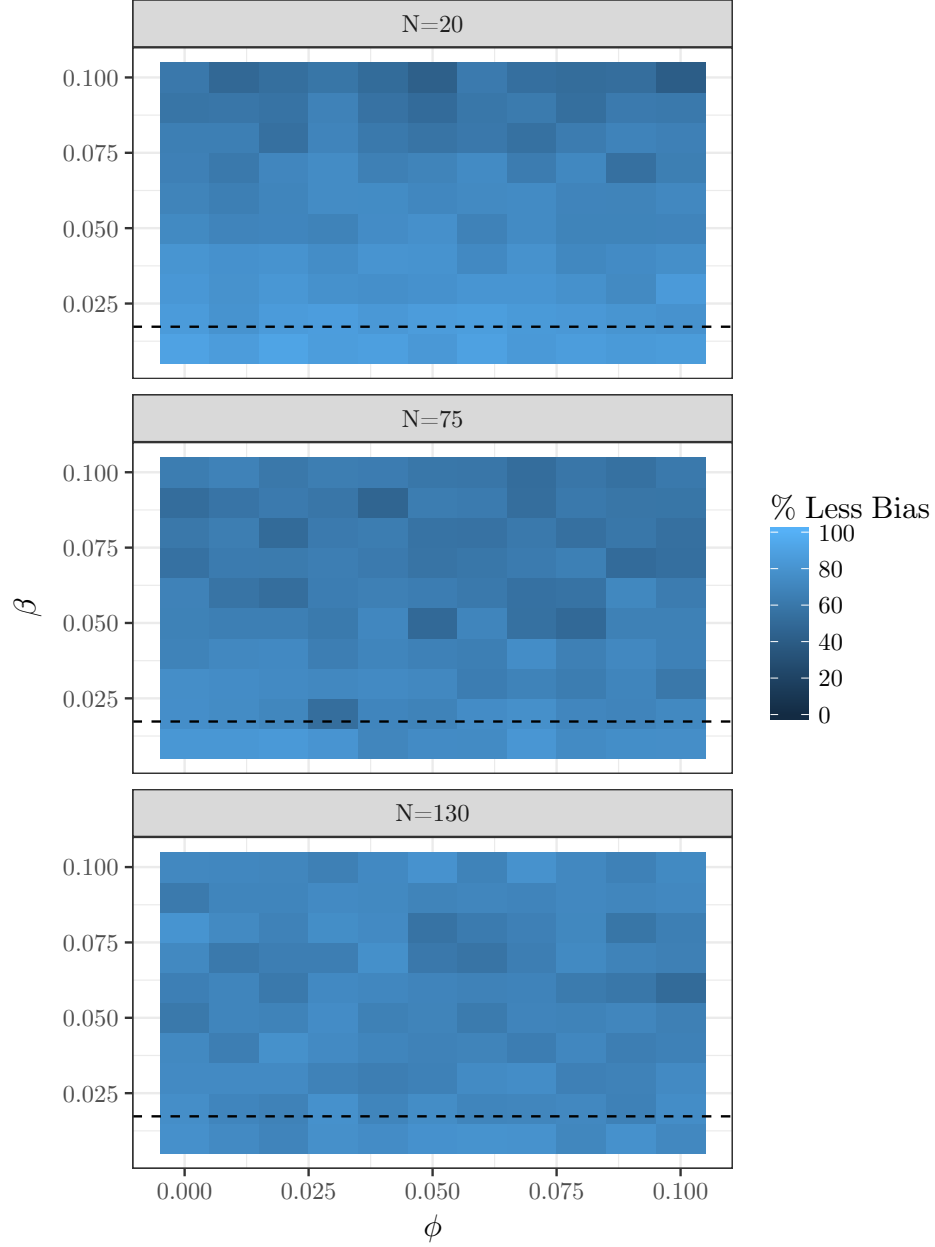
This graph reports results from a variety of interactions. The titles of each panel denote the independent variable and dependent variable under consideration. Each tile illustrates the statistical significance and sign of the effect of that independent variable, conditional on the interacted variable (y-axis) taking a different value from its distribution (along the x-axis). Note that each y-axis-panel combination plots results from a unique model, which is the preferred specification plus the interaction term in question. Section 6 discusses these briefly, and Section E gives more details.

Figure 8: Results from Jackknife Robustness Check



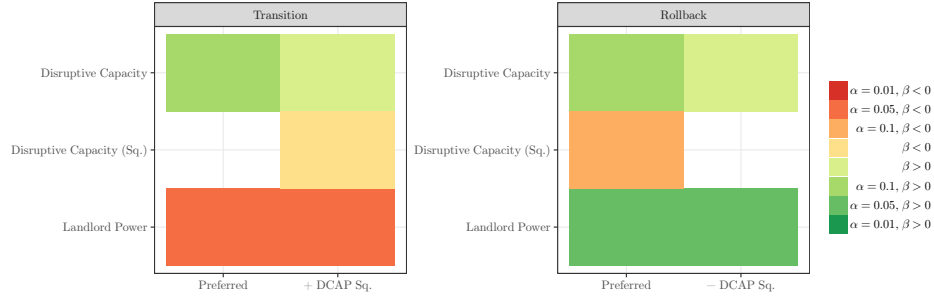
This graph reports the robustness of each of my two key estimates when one country is omitted from the sample. It suggests the main results are not unduly influenced by outlying countries. The estimated impact of disruptive capacity fluctuates slightly, but is always statistically significant at $\alpha = 0.05$ or less.

Figure 9: Relative Gains from LDV vs. Prais-Winsten



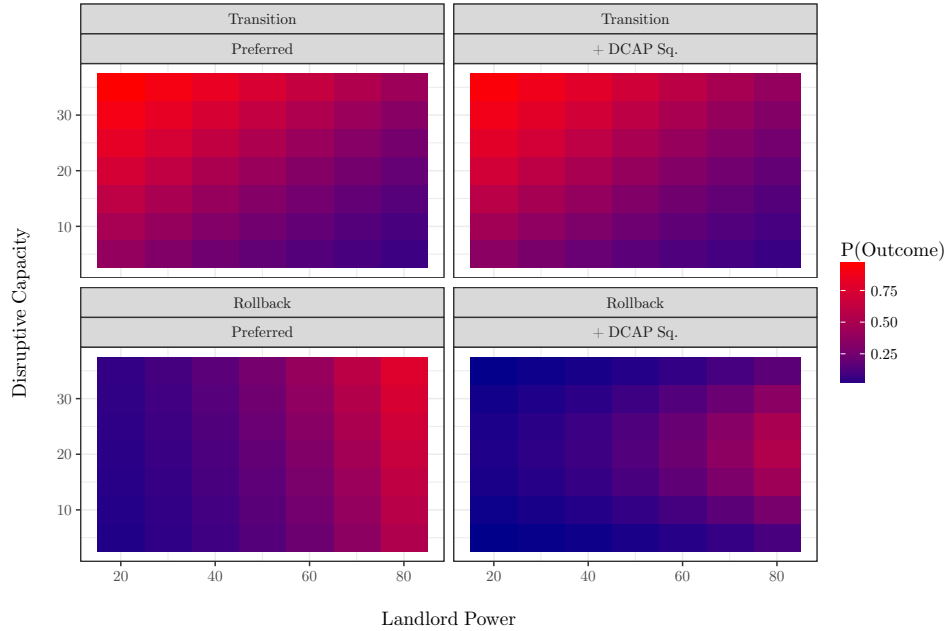
This figure illustrates results from the Monte Carlo simulations I discuss in Section E. I show that models which include a lag of the dependent variable outperform models which do not (i.e. which only account for autocorrelation by adjustment of the assumed error structure). I follow Keele et al. (2006), but focus on the relevant subset of the full parameter space.

Figure 10: Estimates from Models of Transition and Rollback



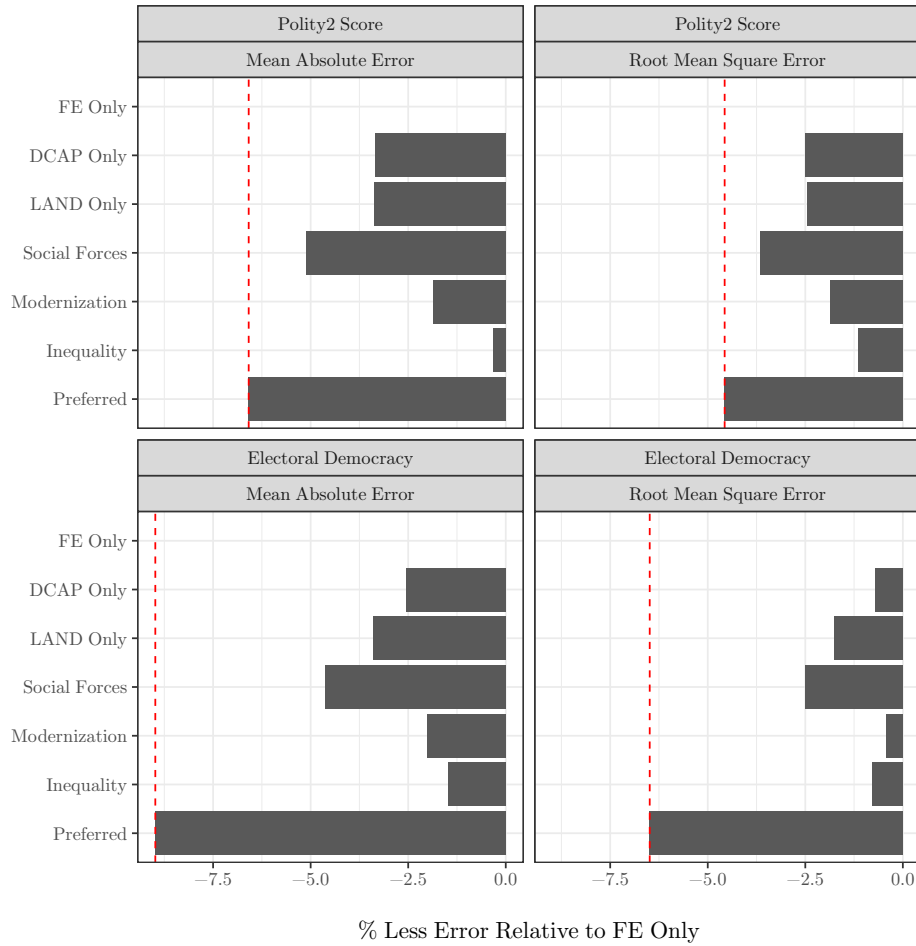
This figure reports coefficients for the two main variables of interest from logistic regressions fit to democratic transitions and authoritarian rollback separately. These refer to the estimated impact of a positive change in the independent variable (on the y-axis) on the probability of the outcome (given in the column heading), across two separate specifications (on the x-axis). The models suggest that landlord power matters as anticipated: negatively associated with transition, and positively associated with rollback. Disruptive capacity matters, though the estimated effect is weakly significant in the transition models, and either absent or quadratic (positive at low levels, negative at higher levels) in the models of authoritarian rollback. See Section E for details of these models and further discussion of these results.

Figure 11: Predicted 10-Year Probabilities of Transition and Rollback



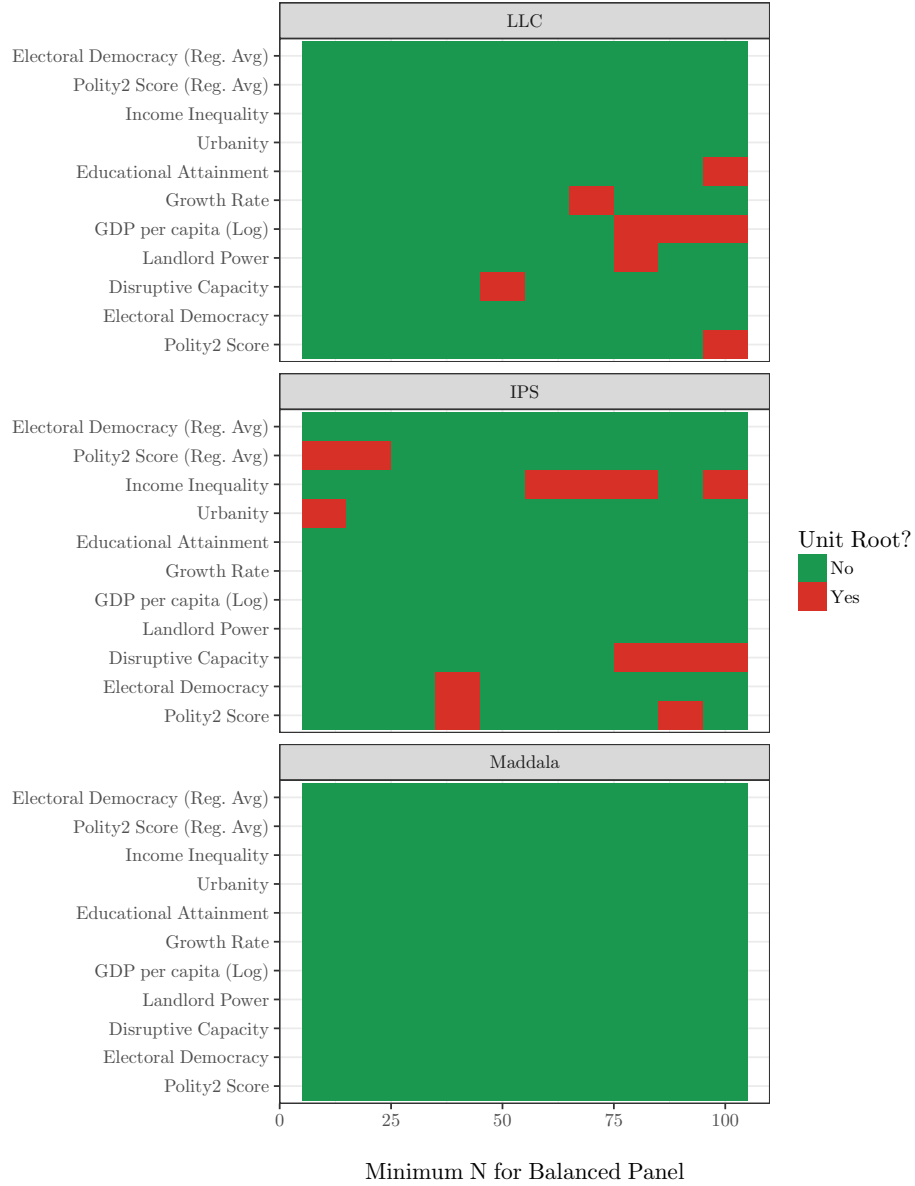
This graph plots the predicted, ten-year probability of democratic transition or authoritarian rollback, based on the binary models of transition and consolidation discussed in Section E. (The probability that the outcome in question occurs in ten years is estimated as the complement of the probability that its complement occurs every year). The x-axis ranges from low to high values of landlord power, and the y-axis ranges from low to high values of disruptive capacity.

Figure 12: Model Fit Statistics



This figure illustrates that my preferred model fits these data better than sparser models. If we take a model that has only year- and country- fixed-effects in it, more complicated models reduce error, however this error is measured. As I discuss in Section E, to obtain these results, I compare simulated trajectories of democratization to observed patterns. These are preferable to conventional measures of fit.

Figure 13: Unit Root Tests



This graph illustrates results from three kinds of unit root tests, run on a constructed and balanced panel of countries observed over years. Section E discusses the construction of this balanced panel in more detail. As the graph shows, my best evidence suggests that the key series I examine in this paper are stationary, and thus justify estimating my main specification in its original levels.

References

- Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2009. "Reevaluating the modernization hypothesis." *Journal of Monetary Economics* 56 (8): 1043–1058.
- Acemoglu, Daron, and James A Robinson. 2006. *Economic origins of dictatorship and democracy*. Cambridge; New York: Cambridge University Press.
- Acemoglu, Daron, James Robinson, Simon Johnson, and Pierre Yared. 2008. "Income and Democracy." *American Economic Review* 98 (3): 808–842.
- Ansell, Ben W., and David J. Samuels. 2014. *Inequality and Democratization: An Elite-Competition Approach*. Cambridge: Cambridge University Press.
- Beck, Nathaniel, David L. Epstein, Simon Jackman, and Sharyn L. O'Halloran. 2001. "Alternative Models of Dynamics in Binary Time-Series-Cross-Section Models: The Example of State Failure."
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42 (4): 1260–1288.
- Carter, David B., and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18 (3): 271–292.
- De Boef, Suzanna, and Luke Keele. 2008. "Taking time seriously." *American Journal of Political Science* 52 (1): 184–200.
- Gelman, Andrew, and Eric Loken. 2013. "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time." *Department of Statistics, Columbia University*.
- Hadri, Kaddour. 2000. "Testing for stationarity in heterogeneous panel data." *The Econometrics Journal* 3 (2): 148–161.
- Haggard, Stephan, and Robert R. Kaufman. 2016. *Dictators and Democrats: Masses, Elites, and Regime Change*. Princeton, NJ: Princeton University Press.
- Im, Kyung So, M. Hashem Pesaran, and Yongcheol Shin. 2003. "Testing for unit roots in heterogeneous panels." *Journal of Econometrics* 115 (1): 53–74.
- Keele, Luke, and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14 (2): 186–205.

- Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu. 2002. "Unit root tests in panel data: asymptotic and finite-sample properties." *Journal of Econometrics* 108 (1): 1–24.
- Lin, Ken-Hou, and Donald Tomaskovic-Devey. 2013. "Financialization and U.S. Income Inequality, 1970–2008." *American Journal of Sociology* 118 (5): 1284–1329.
- Maddala, G. S., and Shaowen Wu. 1999. "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test." *Oxford Bulletin of Economics and Statistics* 61 (S1): 631–652.
- Miller, Michael K. 2012. "Economic Development, Violent Leader Removal, and Democratization." *American Journal of Political Science* 56 (4): 1002–1020.
- Mosley, L., and S. Uno. 2007. "Racing to the bottom or climbing to the top? Economic globalization and collective labor rights." *Comparative Political Studies* 40 (8): 923–948.
- O'Donnell, Guillermo. 1979. *Modernization and Bureaucratic-Authoritarianism: Studies in South American Politics*. Berkeley, CA: University of California Press.
- Przeworski, Adam. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990*. Cambridge: Cambridge University Press.
- Robinson, James A. 2006. "Economic Development and Democracy." *Annual Review of Political Science* 9 (1): 503–527.
- Rudra, Nita. 2002. "Globalization and the Decline of the Welfare State in Less-Developed Countries." *International Organization* 56 (2): 411–445.
- Silver, Beverly J. 2003. *Forces of Labor: Workers' Movements and Globalization Since 1870*. Cambridge: Cambridge Univ Press.