Measuring Backsliding with Observables: Observable-to-Subjective Score Mapping (OSM)

Daniel Weitzel[*] John Gerring[†] Daniel Pemstein[‡] Svend-Erik Skaaning[§]

May 19, 2023

Abstract

Multiple well-known democracy rating projects—including Freedom House, Polity, and V-Dem—have identified apparent global regression in recent years. These measures rely on partly subjective indicators, which could, in principle, suffer from rater bias. For instance, Little and Meng (2023) argue that shared beliefs driven by the current zeitgeist could lead to shared biases that produce the appearance of democratic backsliding in subjectively coded measures. To assess this argument, and the strength of the evidence for global democratic backsliding, we propose an observable-to-subjective score mapping (OSM) methodology that uses only easily observable features of democracy to predict existing indices of democracy. Applying this methodology to three prominent democracy indices, we find evidence of backsliding, but beginning later and not as pronounced as suggested by some of the original indices. Our approach suggests that particularly the Freedom House measure is out of track with the recent patterns in observable indicators and that there has been a stasis or, at most, a modest decline in the average level of democracy.

Word count: 4,692
Keywords: *democracy, backsliding, machine learning*

[*]Department of Political Science, Colorado State University; daniel.weitzel@colostate.edu.
[†]Department of Government, University of Texas at Austin; jgerring@austin.utexas.edu.
[‡]Department of Political Science & Public Policy, North Dakota State University; daniel.pemstein@ndsu.edu.
[§]Department of Political Science, Aarhus University; skaaning@ps.au.dk.

Nearly all extant measures of democracy involve some degree of subjective coding. As is widely recognized, coder judgments may be affected by many factors that introduce error into the coding of a country. This includes personal preferences, political preferences, lack of information, biased sources, varying ideas about how to conceptualize democracy, and data entry mistakes.[1]

Little and Meng (2023) identify one particular bias that could have devastating consequences for our understanding of democracy in the twenty-first century. "Backsliding," they surmise, is part of the zeitgeist, seemingly confirmed by the rise of Trump in the United States, Modi in India, Orban in Hungary, and other populists around the world. This vision of doom is trumpeted by major media, who adopt the backsliding frame in order to explain unfolding events in a readily comprehensible fashion. It is adopted by leaders in the West, who see global forces arrayed on either side of a growing divide – between democracies (the good guys) and autocracies (the bad guys). And it is catnip to a growing industry of democracy scholars and activists whose business is to worry about the fate of democracy. (Arguably, doom-saying enhances the importance and funding available to democracy scholars and activists, and thus serves their interests.)

A more benign interpretation is that democracy indices have become more demanding in their standards. Expert coders, primed to find evidence of backsliding, examine facts on the ground critically, sensing a fundamental threat to democracy in every populist outcry. A high score for democracy is thus harder to achieve in 2020 than it was in 2000 because coders are more attentive to democratic deficits.

Whatever the mechanism(s), it is not hard to see why backsliding might have become a shibboleth of the twenty-first century, one with particular resonance in the West, where democracy indices are headquartered and produced. The hypothetical result is a systematically biased coding of democracy over the past few decades, an apparent downturn that is really the product of a collective miasma or of changes in the way democracy is understood.

To overcome this problem, Little and Meng propose an index of democracy resting solely on observable features relevant to democracy, and hence resistant to errant subjective judgments. Since this index shows little change over the past few decades they argue that there is no evidence of backsliding, at least not on a global scale.

Issues specific to the Little and Meng index are discussed elsewhere in this symposium. Coppedge et al. [cite] demonstrate that the index is prone to problems of conceptualization, indicator selection, aggregation, and coverage, calling into question its utility as a global measure of backsliding. This does not necessarily mean that Little and Meng are wrong about biases toward backsliding in current democracy indices. It does suggest that there may be better ways of addressing the question.

In this study, we adopt an approach that, while also based on observables, may be less susceptible to the problems identified by Coppedge et al. Briefly, we train a random forest model to predict existing indices of democracy using only easily observable features of democracy and a sample limited to years prior to the generally recognized onset of backsliding. We then apply the model to predict scores for the original indices across the entire period, comparing those scores to the ones actually recorded by each index. This approach suggests

[1] See Bush (2017); Cheibub, Gandhi and Vreeland (2010); Giannone (2010); Gründler and Krieger (2016); Skaaning (2018); Steiner (2016); Bowman, Lehoucq and Mahoney (2005); Munck (2009); Weitzel et al. (2023).

that backsliding is real, though beginning later and not as pronounced as the trajectory registered in some democracy indices.

We begin by laying out our methodology, dubbed observable-to-subjective score mapping (OSM) and based on Weitzel et al. (2023). Next, we present the results of our analyses applied to three prominent democracy indices. Section three discusses various robustness tests. Section four considers the missing data problem posed by measuring a latent concept with observables. A short conclusion reflects on what might be learned from this exercise about purported democratic downturns in the twenty-first century.

Methodology

Helpfully, many features relevant to democracy are observable, or relatively so (observability being a matter of degrees). However, it is no mean feat to measure these features in a comprehensive fashion, to select which ones to include in an index, and to arrive at a method of aggregation that has some credibility while preserving the nuance required to discern backsliding.

Nuance is an important consideration in light of the fact that most countries regarded as recent backsliders have not abolished elections, outlawed all opposition parties, or dissolved the legislature. They have, instead, figured out clever ways to undermine the independence of institutions and tilt the electoral playing field in their favor. To capture backsliding it is essential to capture these nuances. Binary indices such as Democracy-Dictatorship (Cheibub, Gandhi and Vreeland 2010) or BMR (Boix, Miller and Rosato 2013) are not sufficient.

Our approach enlists observable-to-subjective score mapping (OSM), a general approach to measurement for situations when both subjectively coded and observable measures for a concept are available (Weitzel et al. 2023). In this instance, we begin with extant indices of democracy, on the presumption that these measurement instruments, which include subjective features, have some prima facie validity – or at least *had* some prima facie validity prior to the current scourge of democratic backsliding.

We focus on three of the most widely used non-dichotomous measures of democracy: the Polyarchy index from Varieties of Democracy (Teorell et al. 2019), the Polity2 index from the Polity IV project (Marshall, Gurr and Jaggers 2015), and the Political Rights and Civil Liberties indicators (combined into a single index by addition) from Freedom House (Freedom House 2021). To assure comparability, we restrict the sample for the following exercise to a common set of 167 polities (see Table SI 5).

Each of the chosen indices forms a target, which we attempt to predict with a wide variety of observable indicators. It is important to be as inclusive as possible in the collection of these indicators so as to avoid arbitrary ("subjective") exclusions that might bias the results. So long as a feature is observable for a broad set of cases and potentially relevant to democracy, it is included in our canvas. A total of 26 indicators as discussed in(Weitzel et al. 2023) are reenlisted here (see Table SI 1.1).

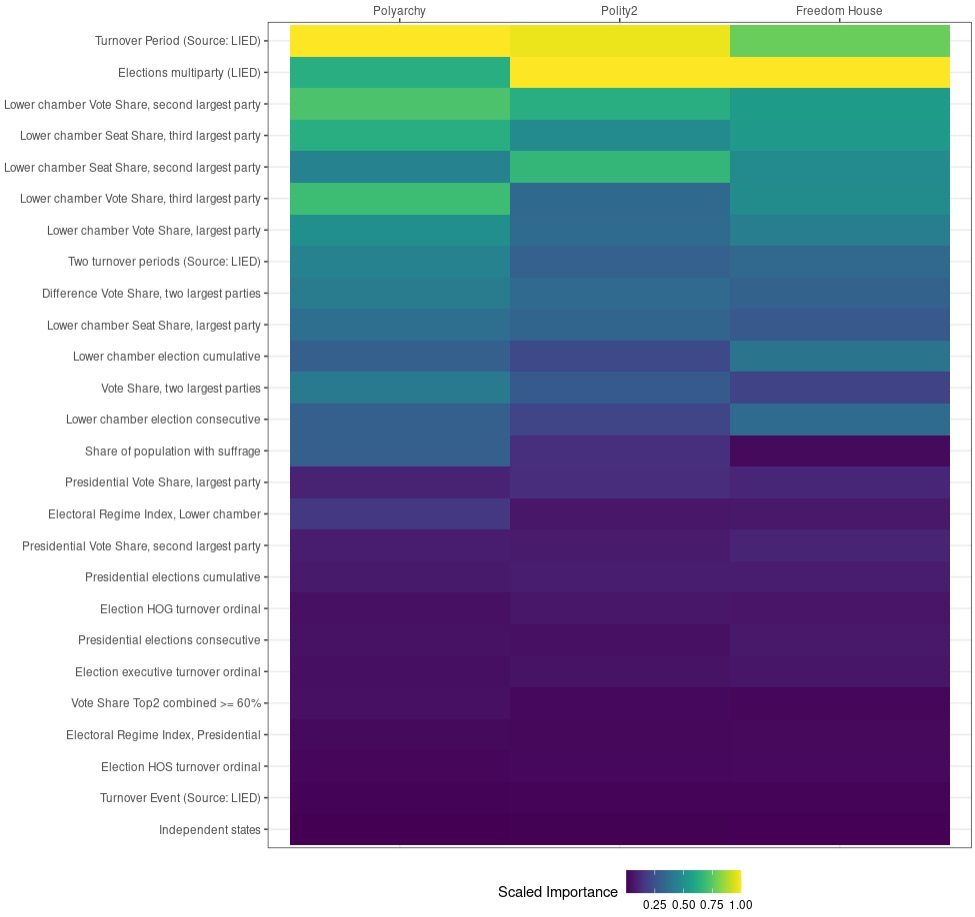
These 26 indicators are treated as predictors in a random forest model in which an existing index of democracy is the outcome to be explained. In this instance, the training set is restricted to the pre-backsliding period, a period when subjective coding was not affected by current expectations of backsliding.

It is an open question when the concept of backsliding, or democratic downturn, first caught on. A Google Ngram, drawing on the Google Books database, shows an uptick in references to "democratic backsliding" around 2010 (see Figure SI 6.1). To avoid any possible overlap, we restrict the training set to years prior to 2000. This training period extends back to 1900 for Polyarchy and Polity2, and back to 1972 for Freedom House, the first year of coding for that index.

We regard this training set as free from the potential bias identified by Little and Meng. Other biases may exist, but these features presumably remain constant through the twentieth and twenty-first centuries and thus do not affect observed trends over time, or are specific to pre-2000 time-periods.

The random forest model assigns weights to each of the 26 variables based on their predictive value. These "importance scores" are depicted in Figure 1. There are marginal differences in the scores assigned to each variable across the three democracy indices. However, importance scores are highly correlated, much like the indices themselves.

Figure 1: Variable importance plot for three democracy indices



Note: Reported is the scaled importance of each variable in the different random forest models. Lighter colors mean a variable is more important in a random forest. Individual variable importance plots can be found in appendix SI 3.

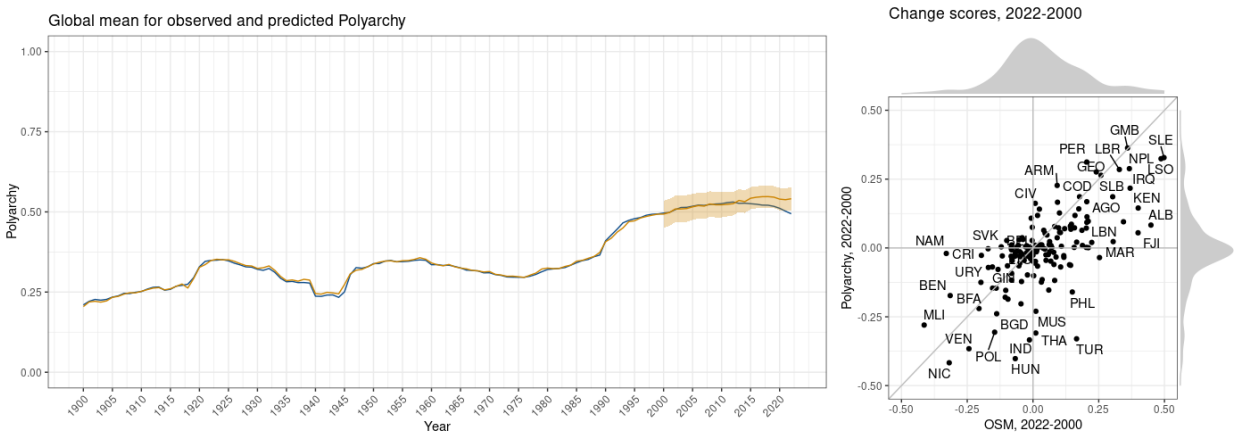
Finally, we use the OSMs to predict values for the post-2000 period. Intuitively, we use the pre-2000 period to learn how to translate the conceptualizations of each index into an aggregation of observable indicators. In other words, we use machine learning to do the hard work of linking measures to concepts.

Post-2000, our model applies these mappings, using only observable indicators as input, purging predictions of any direct human influence, including zeitgeist-driven bias. Out-of-sample predictions extend to 2022 in all cases.[2] If the democratic backsliding reported in subjective indices is due to coder expectations about backsliding or changing standards for democracy — rather than some reality "out there" — these predictions should show no decline.

Results

We begin with the Polyarchy index. Panel (a) in Figure 2 plots the original index and OSM predictions across the entire period of observation, averaging across all 167 countries in our sample (equally weighted).

Figure 2: Polyarchy and OSM Predictions



Note: Panel (a): Polyarchy (dark blue) and OSM predictions for Polyarchy (orange) flanked by 95% confidence intervals from 1900 to 2022. Panel (b): the change in Polyarchy scores (Y axis) against change in OSM predictions (X axis) from 2000-2022, where a positive value indicates an improved democracy score.

OSM predictions track Polyarchy closely, with a small divergence at the very end of the time-period, around 2015, when OSM predictions rise slightly above Polyarchy. Numeric values, recorded in Table 1, show that differences across the two time-series are minute. For example, between 2001 and 2022, the largest difference between Polyarchy and OSM predictions of Polyarchy is 0.048 units on a 0-1 scale. Just a few points at the very end of the time-series fall outside the 95% confidence interval of the mean for the OSM prediction, illustrated by the shaded region in Figure 2.[3]

[2]Data for Polity2 ends in 2018, but our approach relies only on the availability of our observable indicators.

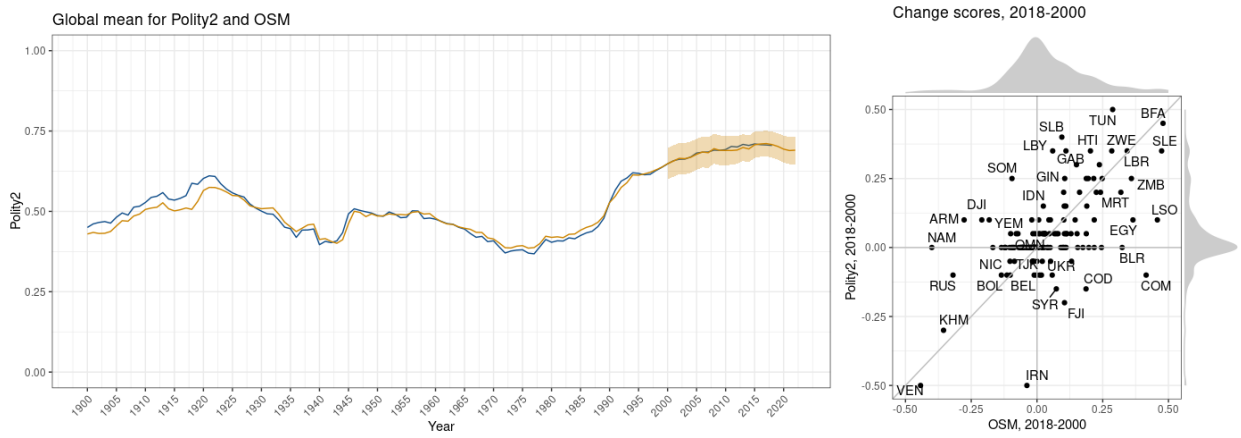
[3]This comparison ignores that Polyarchy is also accompanied by error estimates. In other words, predicted and actual Polyarchy values are even more indistinguishable than the figure implies.

Evidence of a downturn in global democracy can be found in Polyarchy (beginning in 2013) and OSM estimates of Polyarchy (beginning in 2018). However, we must again emphasize the miniscule nature of these changes, especially for the OSM predictions, which do not surpass the confidence interval of prior point estimates.

Leaving global averages aside, we can also look at how particular cases perform during the backsliding period. Panel (b) in Figure 2 focuses on changes registered for specific countries from 2000 to 2022. On the Y axis, we show the change in Polyarchy scores. A score above zero means that the country's democracy score improved; a score below zero means that it deteriorated. The X axis records the same information for OSM predictions.

Most countries lie near the zero point, as shown by the rug plots overlaid along the X and Y axes. Those countries that change scores are situated mostly along the diagonal, demonstrating agreement between Polyarchy and OSM predictions. Several countries fall significantly below the diagonal, indicating that the OSM has a more optimistic view of their trajectory than Polyarchy. This includes Albania, Egypt, Fiji, Hungary, India, Malaysia, Pakistan, Philippines, Thailand, and Turkey. These cases presumably account for the small divergence between the two indicators visible at the end of the time-series in Panel (a) of Figure 2 and in Table 1. These case-rating divergences could reflect excess pessimism in recent Polyarchy ratings, but might also reflect a tendency of changes in observable indicators of democracy to miss difficult-to-observe features available to experts, an issue taken up below.

Figure 3: Polity2 and OSM Predictions



Note: Panel (a): Polity2 (dark blue) and OSM predictions for Polity2 (orange) flanked by 95% confidence intervals from 1900 to 2018 (Polity2) and 2022 (OSM). Panel (b): the change in Polity2 scores (Y axis) against change in OSM predictions (X axis) from 2000 to 2018, where a positive value indicates an improved democracy score.

In Figure 3, we repeat this exercise for Polity2. Panel (a) reveals an even closer fit between the original index and OSM predictions than we saw for Polyarchy. Both curves show some small evidence of backsliding – beginning in 2016 for Polity2 and in 2017 for OSM predictions, as shown in Table 1.

Panel (b) in Figure 3 displays change scores from 2000 to 2018 for all 167 countries as assigned by Polity2 and OSM estimates. Most points again lie close to the center. In a few

Table 1: Global means of democracy indices and OSM predictions

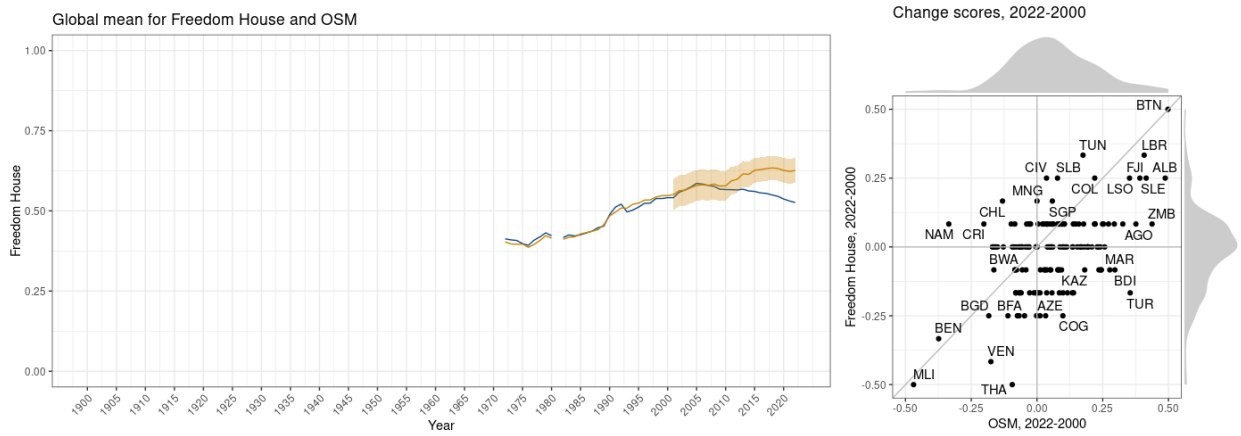
Year	Polyarchy		Polity2		Freedom House	
	Observed	OSM	Observed	OSM	Observed	OSM
2001	0.498	0.498	0.657	0.653	0.541	0.563
2002	0.507	0.502	0.662	0.661	0.558	0.577
2003	0.514	0.507	0.663	0.656	0.565	0.583
2004	0.514	0.508	0.669	0.660	0.575	0.591
2005	0.518	0.518	0.682	0.672	0.586	0.607
2006	0.521	0.523	0.684	0.679	0.584	0.614
2007	0.520	0.527	0.685	0.684	0.580	0.619
2008	0.523	0.533	0.690	0.693	0.577	0.629
2009	0.525	0.530	0.690	0.687	0.568	0.628
2010	0.526	0.528	0.692	0.686	0.567	0.628
2011	0.529	0.529	0.702	0.687	0.566	0.630
2012	0.531	0.533	0.700	0.693	0.565	0.633
2013	0.527	0.538	0.708	0.698	0.568	0.640
2014	0.527	0.540	0.705	0.697	0.562	0.641
2015	0.526	0.548	0.710	0.709	0.561	0.650
2016	0.524	0.550	0.708	0.710	0.556	0.651
2017	0.521	0.553	0.707	0.709	0.555	0.652
2018	0.521	0.550	0.705	0.705	0.550	0.651
2019	0.517	0.548		0.701	0.546	0.650
2020	0.511	0.544		0.694	0.537	0.644
2021	0.502	0.541		0.689	0.531	0.640
2022	0.494	0.542		0.688	0.526	0.641

Note: Global means of three democracy indices (rescaled from 0 to 1) along with out-of-sample OSM predictions for those indices.

instances the OSM offers different assessments. Polity2 is more pessimistic than the OSM about the twenty-first century trajectories of Comoros, Democratic Republic of the Congo, Fiji, and Iran, for example.

Figure 4 completes our exercise, this time focused on the Freedom House index. Panel (a) shows that Freedom House registers a fairly sharp downturn beginning in 2006. Meanwhile, the OSM predictions continue ascending through 2017, after which there is a modest downturn. Unlike for Polyarchy and Polity2, out-of-sample OSM predictions for Freedom House diverge dramatically, with recent observations falling well outside the confidence intervals.

Figure 4: Freedom House and OSM Predictions



Note: Panel (a): Freedom House (dark blue) and OSM predictions for Freedom House (orange) flanked by 95% confidence intervals from 1972 to 2022. Panel (b): the change in Freedom House scores (Y axis) against change in OSM predictions (X axis) from 2000-2021, where a positive value indicates an improved democracy score. The gap from 1981 to 1982 is a gap in Freedom House data collection.

These divergences are also notable in Panel (b) of Figure 4, where the OSM has a decidedly more optimistic view of regime changes in Turkey, the Republic of Congo, and Burundi, and a more pessimistic view of developments in Costa Rica and Namibia.

One might regard the divergence between Freedom House and the OSM model as a failing of our modeling approach, especially as certain features of the Freedom House index appear to complicate the task of making out-of-sample predictions beyond the observed time-series. First, the index is sluggish, registering few changes through time relative to Polyarchy and Polity2 (see Table SI 10.3). Second, because the Freedom House index begins in 1972 we do not have an extended sample upon which to train the random forest model.

However, when the same sample restriction is imposed on Polyarchy and Polity2 we observe only a modest attenuation in fit, so the shortness of the sample cannot be the whole story. Moreover, results displayed in Panel (a) of Figure 4 show that the OSM is quite proficient in predicting the first several years of Freedom House, out of sample. Large differences appear only after 2005.

One explanation for the divergence between Freedom House and the OSM is that coding principles changed around the 2006 edition of Freedom House, leading to a fundamentally different data generating process that the OSM model could not — and, by design, should not — replicate. As it happens, 2006 is the first year that Freedom House publicly released sub-category scores for its extensive questionnaire. At the same time, the number of coders (analysts) increased steeply (from 14 to 23), after which this number continues to grow, reaching a total of 128 for the 2023 report. Finally, the 2006 edition introduced a rewording of several survey questions and of the coding guidelines.

It seems quite possible that one or more of these modifications account for the divergence between the ratings submitted by Freedom House and those predicted out of sample by the OSM. This interpretation seems to be corroborated by other democracy indices such as Polyarchy and Polity2, where global downturns do not appear until the second decade of the

twenty-first century (see Figure SI 9.1).

Robustness Tests

As with any complex measurement exercises, there are many moving parts to this approach. Weitzel et al. (2023) discuss the robustness of OSMs with respect to the measurement of democracy. Here, we focus on issues bearing directly on out-of-sample predictions lying outside the observed time-series.

A critical question concerns the cutoff point separating the training set from the out-of-sample test set. We chose 2000 for our benchmark OSM model because it falls well before the period usually identified as prone to backsliding. In additional tests, we set the cutoff at 2005 and 2010, respectively. Results displayed in appendices SI 7 and SI 8 are nearly identical to those generated from the benchmark cutoff (displayed in previous figures and tables). We conclude that the cutoff has minimal influence on our findings.

A second issue concerns the selection of indicators for the OSM model. To test robustness, we exclude influential predictors (those with high importance scores as shown in Figure 1) seriatim. With each exclusion, the process outlined above is repeated. Results of each iteration are very close to those reported in the benchmark model (with the full set of 26 variables).

A third issue concerns the ability of the OSM model to predict out-of-sample observations beyond the observed time-series when a long-term trend changes, e.g., in a downward direction. Note that divergences appear at about the point when the Polyarchy and Freedom House trends turn downward.

We expect subtle changes in the quality of democracy to register in the OSM with a lag since most of our observable indicators are associated with elections. If the ruling party shuts down the main opposition press and this enhances the party's vote in the next election it will not be registered in the OSM model until that election year. However, the OSM should catch up with reality fairly quickly since elections occur regularly. (In the event that elections are canceled or indefinitely postponed this is registered by a change in coding for the electoral regime variables, and thus should appear immediately in the OSM.)

To see whether the OSM is able to track big changes in the level of global democracy prior to the twenty-first century we conduct several tests in which the model is trained on a period when democracy's ascent is highly trended and then tested on out-of-sample observations when that trend changes. We find that notable downward trends in the interwar period and the 1970s reported by Polyarchy and Polity2 are well-predicted by the model (see SI 12, SI 13). This suggests that divergences in the twenty-first century are not the product of changes to the overall trend.

The Missing Data Problem

Before concluding, we must consider an alternate explanation for the divergence between the OSM and Freedom House (and, to a lesser extent, Polyarchy).

Perhaps the OSM model offers a more optimistic picture of global democracy in recent years because it does not capture features of regimes that are not directly observable, and

hence not included in our collection of indicators. Freedom of speech, for example, is famously difficult to measure with observable indicators and is often cited as a deficit in backsliding regimes.

For missing data to impair OSM estimates of the global trend those missing features would have to become more prominent or more important (in determining regime type) over time, while measurable features of democracy remain constant. Let us consider this scenario carefully.

Suppose that media outlets throughout the world experienced less independence after 2005 than previously. If this was the only change in democracy over the twenty-first century it might account for why Freedom House scores attenuate while OSM estimates of those scores, trained on data prior to 2000 and bereft of any direct measure of media independence, do not.

However, if the independence of the media was seriously compromised one would expect this to impact observable features of the quality of democracy. In particular, restrictions on free speech should boost the electoral performance of the incumbent party (which would be registered in our election-related variables) or they would be accompanied by a formal proroguing of parliament and cessation of elections (registered in our electoral regime variables). If violations of media independence are *not* correlated with these and other observable features of democracy one may question how consequential those violations are. Perhaps citizens heavily discount the veracity of official channels and have access to alternate web-based news sources.

The logic of measuring democracy with observables does not rest on the assumption that everything relevant to democracy is observable. It rests on the assumption that most things relevant to democracy are correlated with features that are observable.

To conclude, missing data of a subjective nature may explain the small divergence between Polyarchy and OSM estimates of Polyarchy that appears in the past decade (see Figure 2). But it is unlikely to account for the much larger divergence between Freedom House and OSM estimates that appear after 2005 (see Figure 4).

Backsliding Reconsidered

In this study, we assess the veracity of recent claims of democratic backsliding using machine learning. We train a random forest model, informed by observable indicators of democracy, on target data from three prominent composite indices of democracy, during a period prior to the alleged backsliding. This "OSM" model issues predictions for the out-of-sample period, when backsliding is thought to have taken hold.

Applying this methodology, we find evidence of backsliding using all three indices. However, the OSM estimates are not always in perfect accord with the original indices. OSM estimates of Polity2 follow the original index scores closely. OSM estimates of Polyarchy suggest that backsliding on a global scale may have started later, and may be somewhat more modest than is suggested by Polyarchy. However, uncertainty intervals around the OSM predictions contain almost all of Polyarchy values so one would not want to over-interpret small differences toward the end of the time-series.

In the case of Freedom House, we find wide divergence between the original index and

OSM estimates. We hypothesize that this divergence is indicative of changes to the coding process in Freedom House, as discussed.[4]

It is important to stress that the machine learning approach, as employed here, does not assess validity. Rather, it assesses consistency through time. If the data generating process changes this will produce divergence between OSM estimates and the original index. One might view FH II (post-2005) as an improvement over FH I (pre-2005). This we cannot determine. What we can suggest is that Freedom House is probably not a very reliable instrument for measuring changes in democracy over time.

What, then, shall we conclude about global democratic backsliding?

Our analysis offers some support for a broad version of the Little and Meng hypothesis. Some indices, and most especially Freedom House, may be subject to time-specific measurement bias. At the very least, there is variability through time in how their assessments translate facts on the ground.

At the same time, and in contrast to Little and Meng, we do find evidence of a slight global downturn in democracy since 2018 using only observable indicators trained on data from the pre-2000 period. In this respect, our exercise validates conventional wisdom.

We would emphasize the slightness and shortness of this declension. Even so, it is newsworthy that the advance of democracy that began two centuries ago, with only a few large-scale reversals, has halted. One might conclude that *stasis*, rather than backsliding, should be the headline of this symposium.

[4]This impression is reinforced when we look at previous eras of Freedom House coding. After training the OSM on data prior to 1988, out-of-sample OSM estimates sharply diverge from FH in the early 1990s (see SI 11.1). This was also the period in which Raymond Gastil, the original coder – and sole coder for many years – of the Freedom House index, left the project.

References

Boix, Carles, Michael Miller and Sebastian Rosato. 2013. "A complete data set of political regimes, 1800–2007." *Comparative Political Studies* 46(12):1523–1554.

Bowman, Kirk, Fabrice Lehoucq and James Mahoney. 2005. "Measuring political democracy: Case expertise, data adequacy, and Central America." *Comparative Political Studies* 38(8):939–970.

Bush, Sarah Sunn. 2017. "The politics of rating freedom: Ideological affinity, private authority, and the Freedom in the World ratings." *Perspectives on Politics* 15(3):711–731.

Cheibub, José Antonio, Jennifer Gandhi and James Raymond Vreeland. 2010. "Democracy and dictatorship revisited." *Public choice* pp. 67–101.

Freedom House. 2021. "Freedom in the World 2020 Methodology Introduction-History of Freedom in the World.".

Giannone, Diego. 2010. "Political and ideological aspects in the measurement of democracy: the Freedom House case." *Democratization* 17(1):68–97.

Gründler, Klaus and Tommy Krieger. 2016. "Democracy and growth: Evidence from a machine learning indicator." *European Journal of Political Economy* 45:85–107.

Little, Andrew and Anne Meng. 2023. "Subjective and Objective Measurement of Democratic Backsliding.".

Marshall, Monte G, Ted Robert Gurr and Keith Jaggers. 2015. "Dataset User's Manual. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2015." *Center for Systemic Peace* .

Munck, Gerardo L. 2009. *Measuring democracy: A bridge between scholarship and politics*. JHU Press.

Skaaning, Svend-Erik. 2018. "Different types of data and the validity of democracy measures." *Politics and Governance* 6(1):105–116.

Steiner, Nils D. 2016. "Comparing Freedom House democracy scores to alternative indices and testing for political bias: Are US allies rated as more democratic by Freedom House?" *Journal of Comparative Policy Analysis: Research and Practice* 18(4):329–349.

Teorell, Jan, Michael Coppedge, Staffan Lindberg and Svend-Erik Skaaning. 2019. "Measuring polyarchy across the globe, 1900–2017." *Studies in Comparative International Development* 54:71–95.

Weitzel, Daniel, John Gerring, Daniel Pemstein and Svend-Erik Skaaning. 2023. "Measuring Electoral Democracy with Observables.".

SUPPORTING INFORMATION

SI 1 Variables

Table SI 1.1: Outcomes and predictors

Democracy Indicators	
Electoral democracy index (D)	*v2x_polyarchy*
Freedom House, combined	*e_fh_combined*
Polity revised combined score (E)	*e_polity2*
Predictors	
Lower chamber election vote share of largest vote-getter (A)	*v2ellovtlg*
Lower chamber election vote share of second-largest vote-getter (A)	*v2ellovtsm*
Lower chamber election vote share of third-largest vote-getter (A)	*v2ellovttm*
Lower chamber election seat share won by largest party (A)	*v2ellostsl*
Lower chamber election seat share won by second largest party (A)	*v2ellostss*
Lower chamber election seat share won by third largest party (A)	*v2ellostts*
Presidential election vote share of largest vote-getter (A)	*v2elvotlrg*
Presidential election vote share of second-largest vote-getter (A)	*v2elvotsml*
Executive electoral regime index (A)	*v2xex_elecreg*
Legislative electoral regime index (A)	*v2xlg_elecreg*
Elections multiparty (LIED)	*multi_party_elections*
Share of population with suffrage (D)	*v2x_suffr*
Dummy for legislative elections	*v2eltype_legislative*
Dummy for presidential elections	*v2eltype_presidential*
Difference in vote share of top two parties	*top2_difference*
Combined vote share of top two parties	*top2_combined*
Top two parties have vote share larger than 59.99	*top2_monopoly*
Legislative elections, consecutive	*v2ellocons*
Legislative elections, cumulative	*v2ellocumul*
Presidential elections, consecutive	*v2ellocons*
Presidential elections, cumulative	*v2elprescumul*
Head of government turnover	*v2elturnhog*
Head of state turnover	*v2elturnhos*
Executive turnover	*v2eltvrexo*
Turnover period (LIED)	*turnover_period*
Turnover event (LIED)	*turnover_event*
Two turnover period (LIED)	*two_turnover_period*

SI 2 Missing Data Plot

Figure SI 2.1: Missing Data Plot

SI 3 Variable Importance Plot

Figure SI 3.1: Variable Importance Plot for Polyarchy

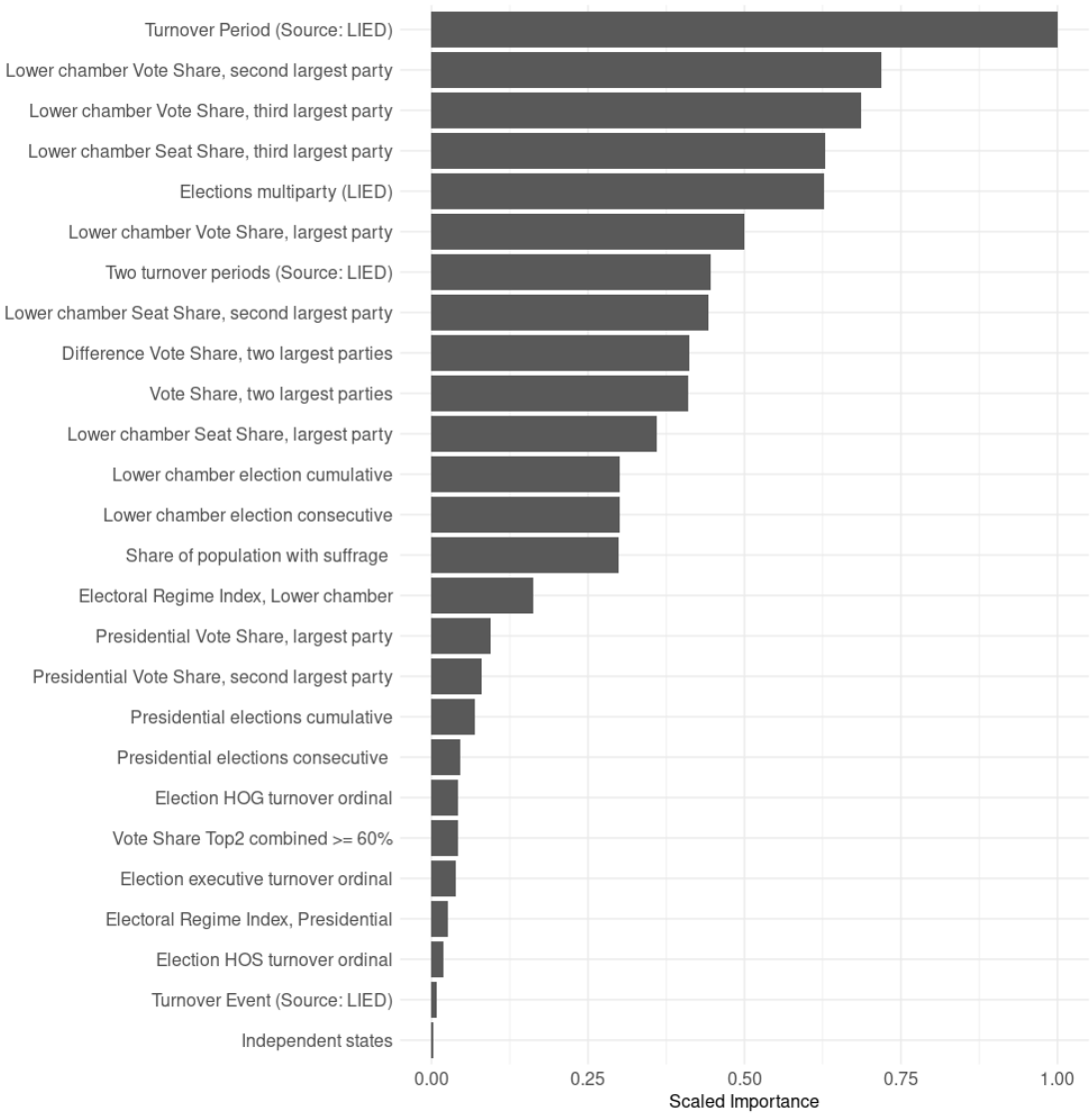


Figure SI 3.2: Variable Importance Plot for Polity2

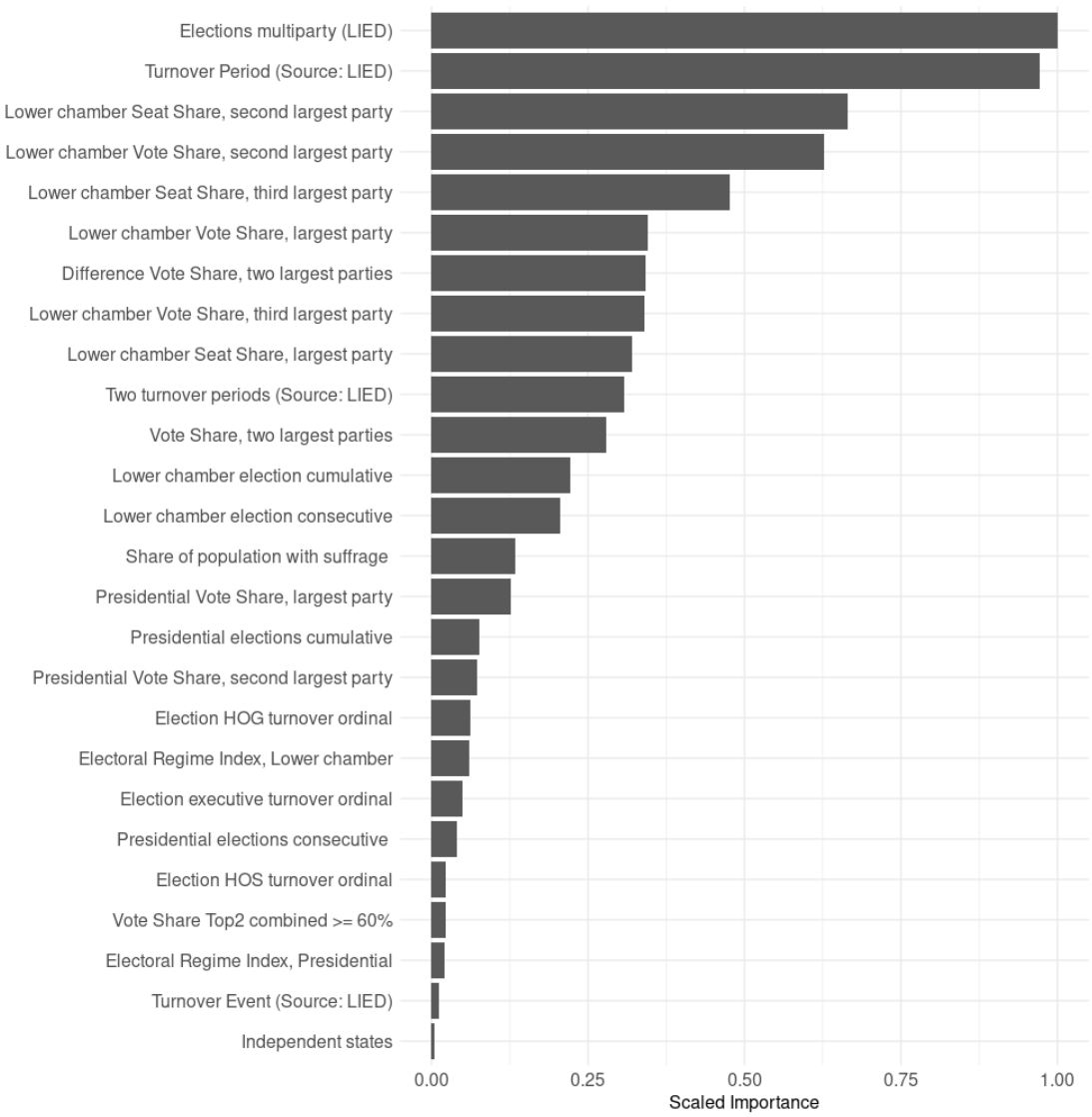
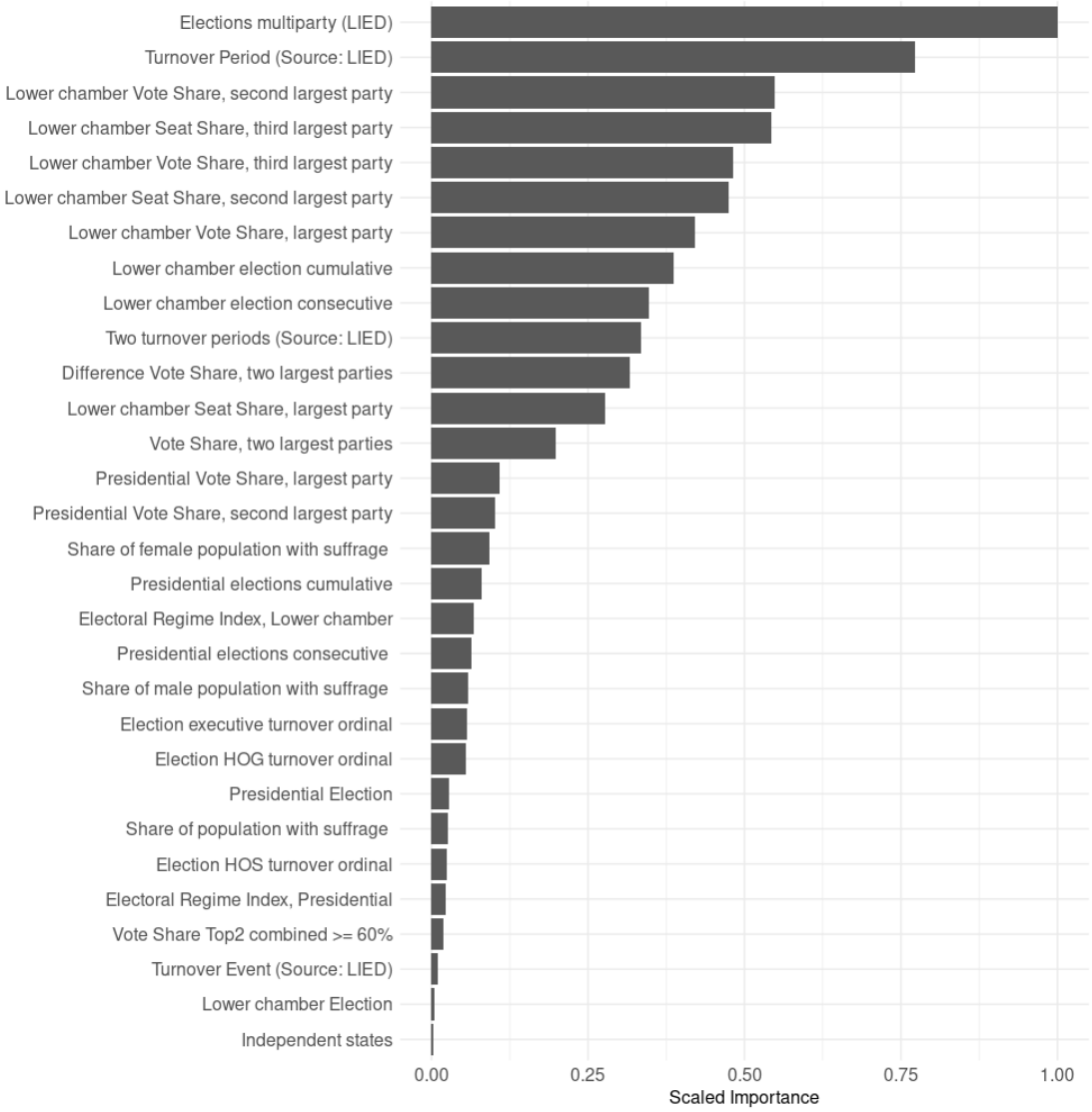


Figure SI 3.3: Variable Importance Plot for Freedom House



SI 4 Histogram and Density Plots for out of sample prediction

Figure SI 4.1: Histogram and Density for Polyarchy

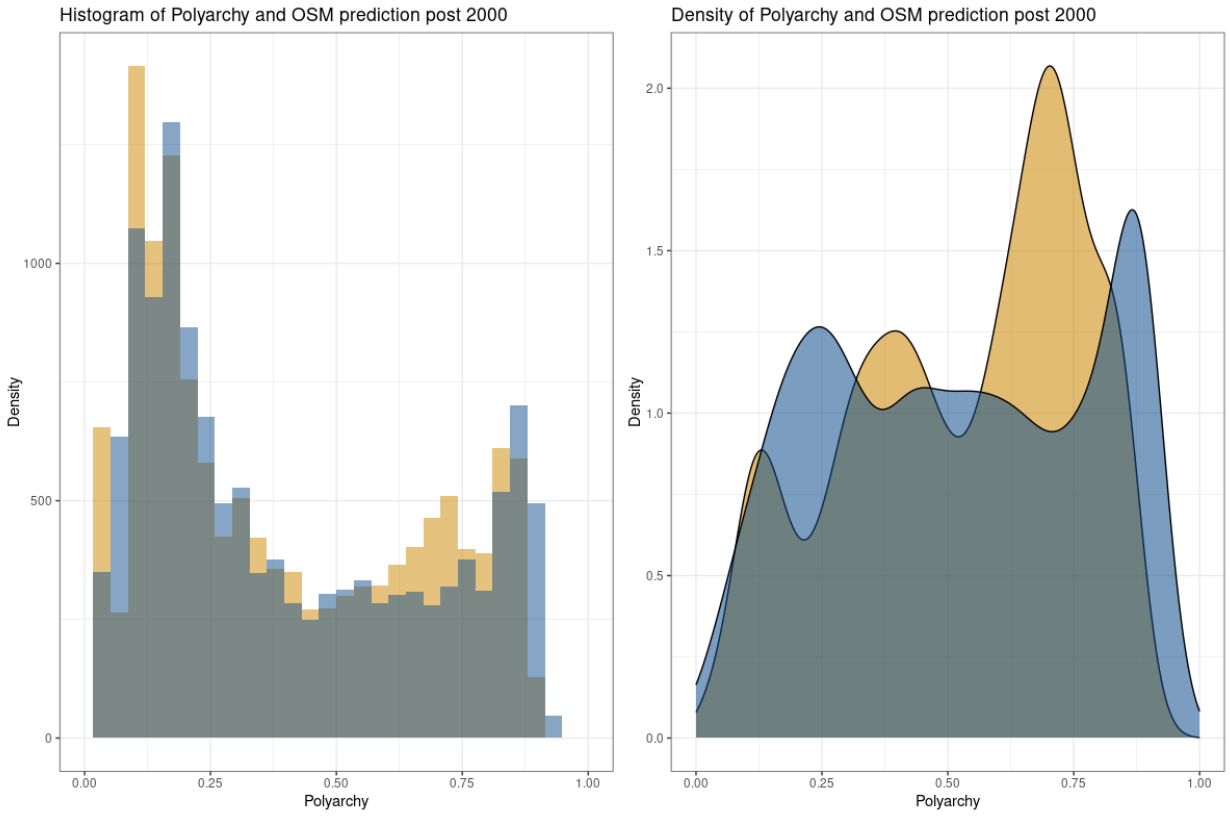


Figure SI 4.2: Histogram and Density Plot for Polity2

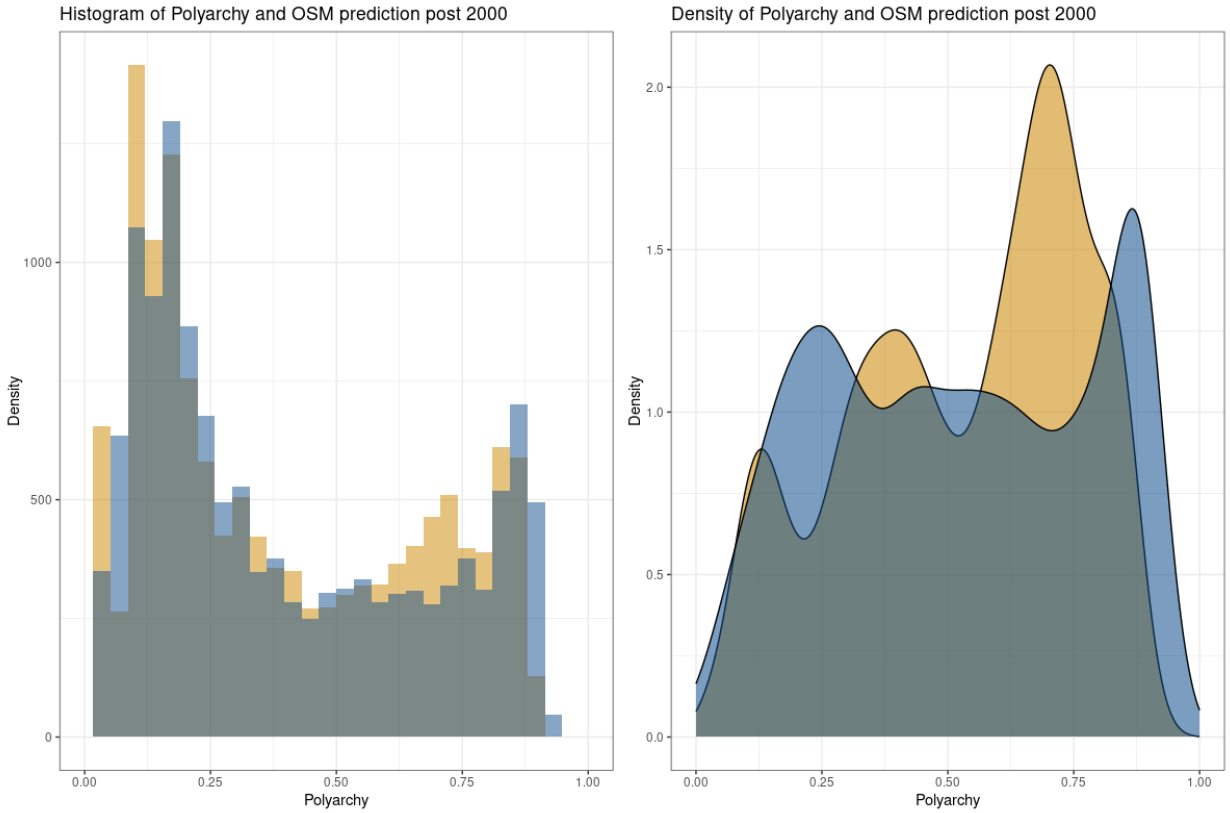
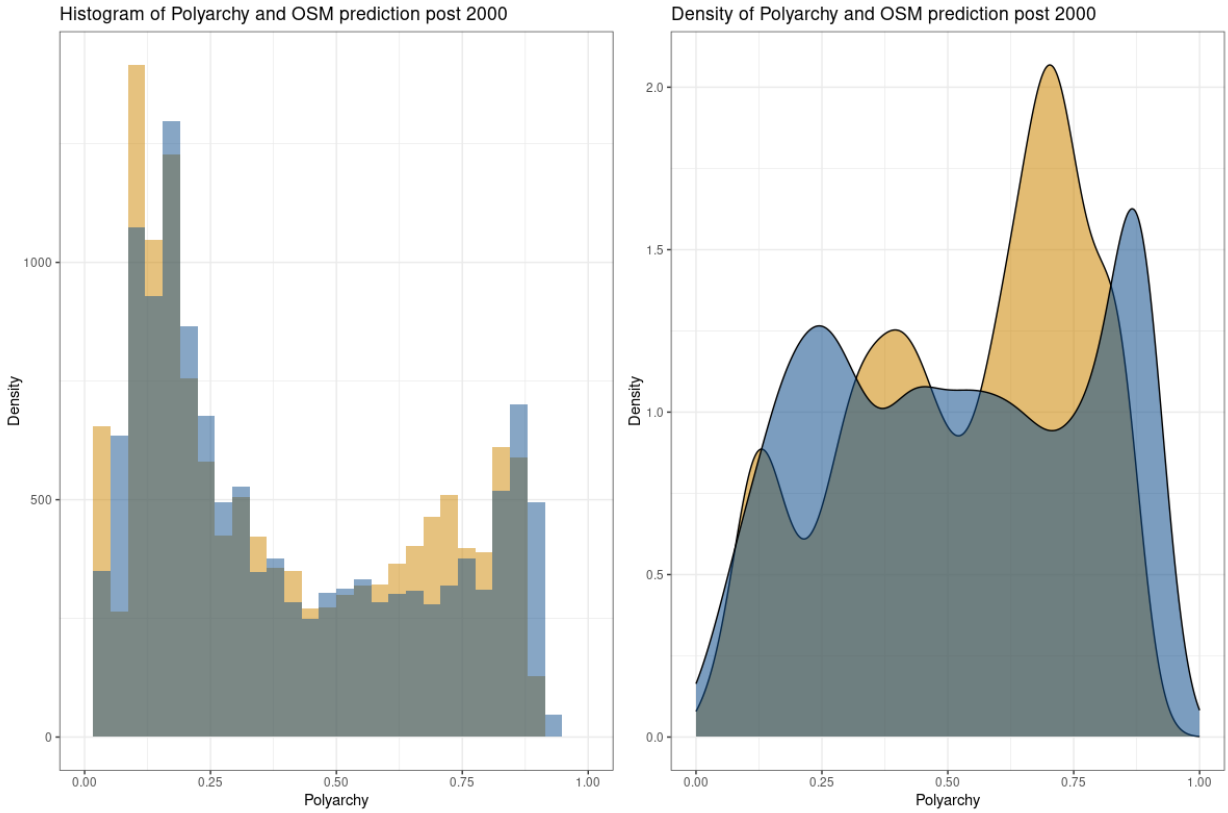


Figure SI 4.3: Histogram and Density Plot for Freedom House



SI 5 Polities

Countries			
AFG	Afghanistan	LAO	Laos
AGO	Angola	LBN	Lebanon
ALB	Albania	LBR	Liberia
ARE	United Arab Emirates	LBY	Libya
ARG	Argentina	LKA	Sri Lanka
ARM	Armenia	LSO	Lesotho
AUS	Australia	LTU	Lithuania
AUT	Austria	LUX	Luxembourg
AZE	Azerbaijan	LVA	Latvia
BDI	Burundi	MAR	Morocco
BEL	Belgium	MDA	Moldova
BEN	Benin	MDG	Madagascar
BFA	Burkina Faso	MEX	Mexico
BGD	Bangladesh	MKD	North Macedonia
BGR	Bulgaria	MLI	Mali
BHR	Bahrain	MMR	Burma/Myanmar
BIH	Bosnia and Herzegovina	MNE	Montenegro
BLR	Belarus	MNG	Mongolia
BOL	Bolivia	MOZ	Mozambique
BRA	Brazil	MRT	Mauritania
BTN	Bhutan	MUS	Mauritius
BWA	Botswana	MWI	Malawi
CAF	Central African Republic	MYS	Malaysia
CAN	Canada	NAM	Namibia
CHE	Switzerland	NER	Niger
CHL	Chile	NGA	Nigeria
CHN	China	NIC	Nicaragua
CIV	Ivory Coast	NLD	Netherlands
CMR	Cameroon	NOR	Norway
COD	Democratic Republic of the Congo	NPL	Nepal
COG	Republic of the Congo	NZL	New Zealand
COL	Colombia	OMN	Oman
COM	Comoros	PAK	Pakistan
CPV	Cape Verde	PAN	Panama
CRI	Costa Rica	PER	Peru
CUB	Cuba	PHL	Philippines
CYP	Cyprus	PNG	Papua New Guinea
CZE	Czechia	POL	Poland
DDR	German Democratic Republic	PRK	North Korea
DEU	Germany	PRT	Portugal

Continued on next page...

Table SI 5.2 (Continued)

Countries			
DJI	Djibouti	PRY	Paraguay
DNK	Denmark	QAT	Qatar
DOM	Dominican Republic	ROU	Romania
DZA	Algeria	RUS	Russia
ECU	Ecuador	RWA	Rwanda
EGY	Egypt	SAU	Saudi Arabia
ERI	Eritrea	SDN	Sudan
ESP	Spain	SEN	Senegal
EST	Estonia	SGP	Singapore
ETH	Ethiopia	SLB	Solomon Islands
FIN	Finland	SLE	Sierra Leone
FJI	Fiji	SLV	El Salvador
FRA	France	SOM	Somalia
GAB	Gabon	SRB	Serbia
GBR	United Kingdom	SSD	South Sudan
GEO	Georgia	SUR	Suriname
GHA	Ghana	SVK	Slovakia
GIN	Guinea	SVN	Slovenia
GMB	The Gambia	SWE	Sweden
GNB	Guinea-Bissau	SWZ	Eswatini
GNQ	Equatorial Guinea	SYR	Syria
GRC	Greece	TCD	Chad
GTM	Guatemala	TGO	Togo
GUY	Guyana	THA	Thailand
HND	Honduras	TJK	Tajikistan
HRV	Croatia	TKM	Turkmenistan
HTI	Haiti	TLS	Timor-Leste
HUN	Hungary	TTO	Trinidad and Tobago
IDN	Indonesia	TUN	Tunisia
IND	India	TUR	Turkey
IRL	Ireland	TWN	Taiwan
IRN	Iran	TZA	Tanzania
IRQ	Iraq	UGA	Uganda
ISR	Israel	UKR	Ukraine
ITA	Italy	URY	Uruguay
JAM	Jamaica	USA	United States of America
JOR	Jordan	UZB	Uzbekistan
JPN	Japan	VDR	Republic of Vietnam
KAZ	Kazakhstan	VEN	Venezuela
KEN	Kenya	VNM	Vietnam
KGZ	Kyrgyzstan	XKX	Kosovo
KHM	Cambodia	YEM	Yemen

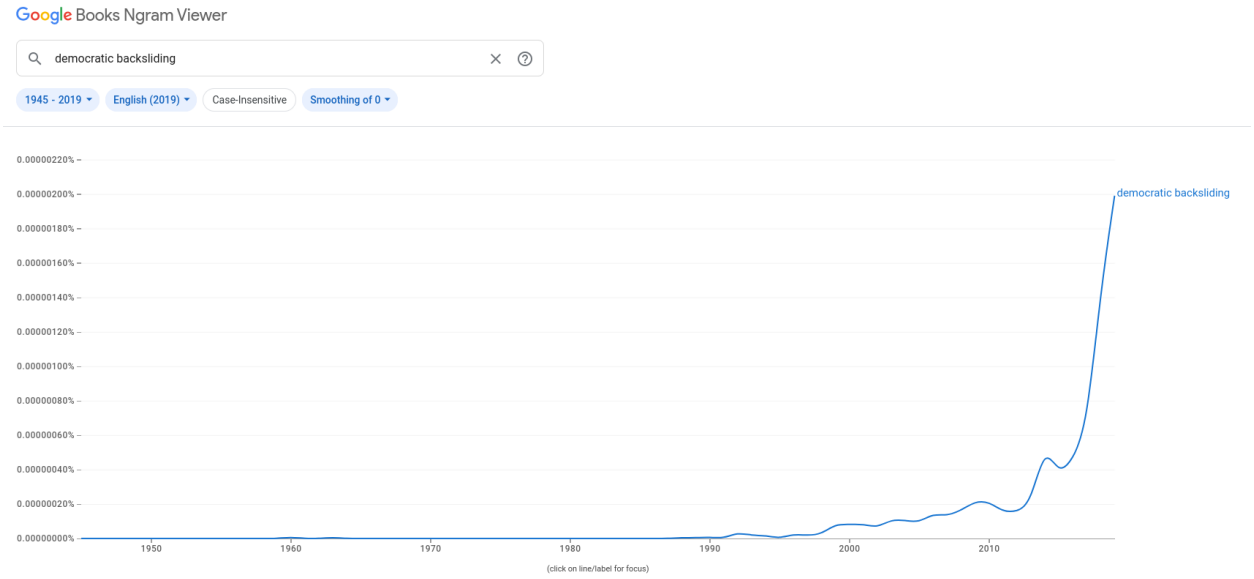
Continued on next page...

Table SI 5.2 (Continued)

Countries			
KOR	South Korea	YMD	South Yemen
KWT	Kuwait		

SI 6 Google Ngram for Democratic Backsliding

Figure SI 6.1: Google Ngram for Democratic Backsliding



SI 7 Cut off at 2005

Figure SI 7.1: Cut off at 2005 for Polyarchy

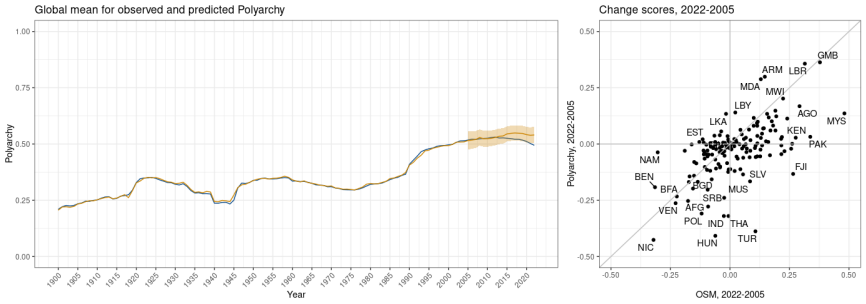


Figure SI 7.2: Cut off at 2005 for Polity2

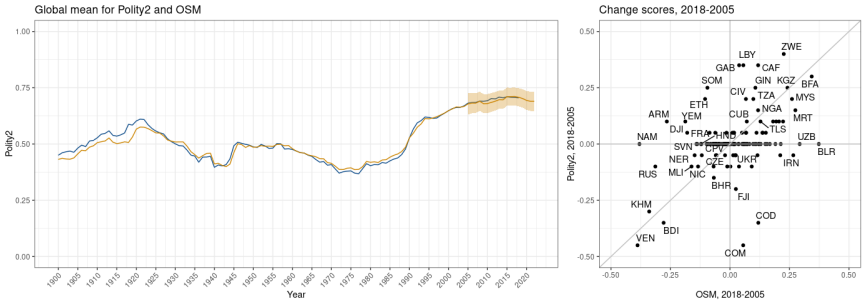
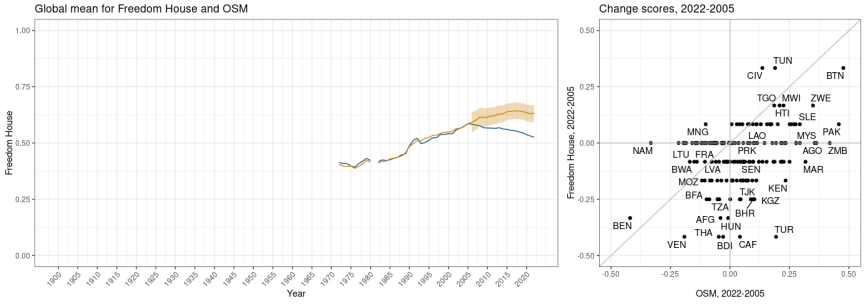


Figure SI 7.3: Cut off at 2005 for Freedom House



SI 8 Cut off at 2010

Figure SI 8.1: Cut off at 2010 for Polyarchy

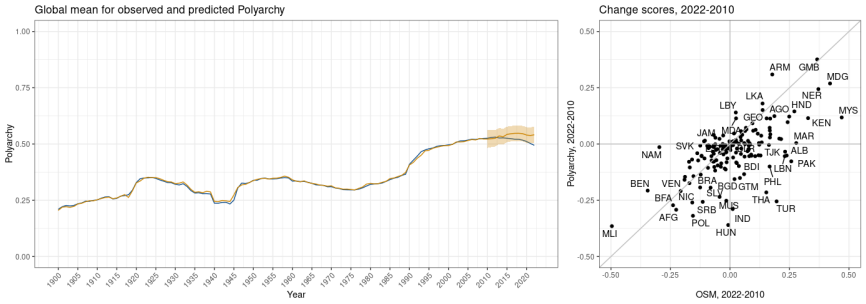


Figure SI 8.2: Cut off at 2010 for Polity2

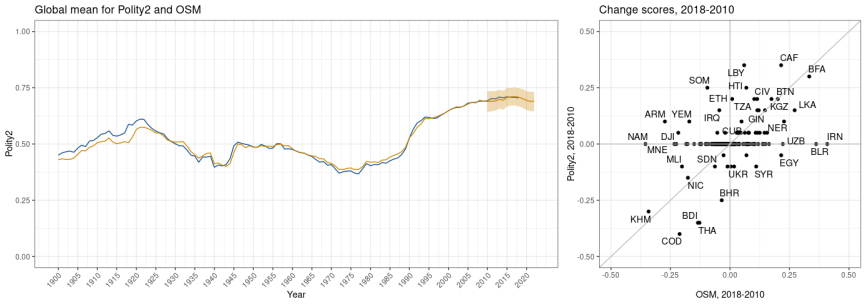
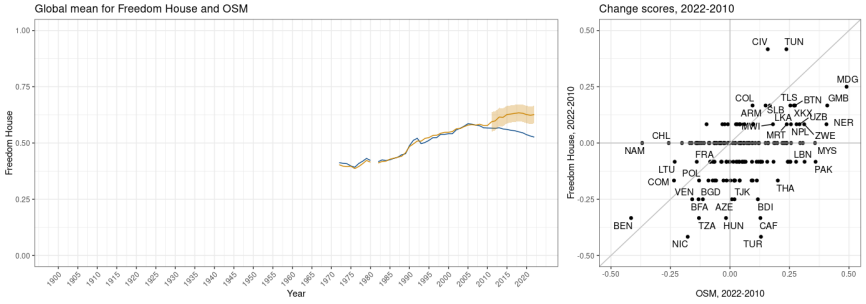
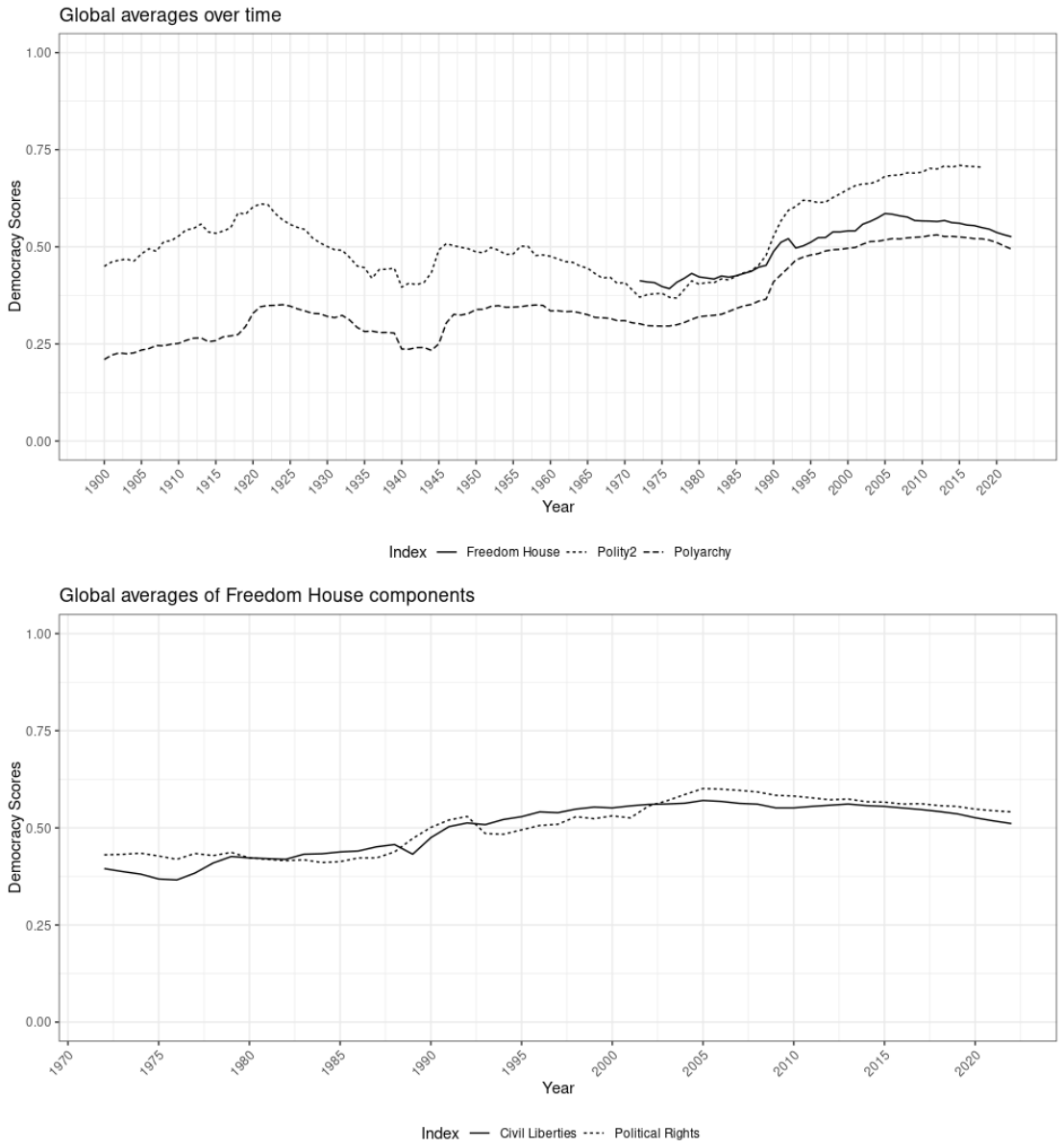


Figure SI 8.3: Cut off at 2010 for Freedom House



SI 9 Global means of democracy indicators

Figure SI 9.1: Global means of democracy indicators



SI 10 Annual changes in democracy scores of countries

Figure SI 10.1: Annual changes in democracy scores of countries

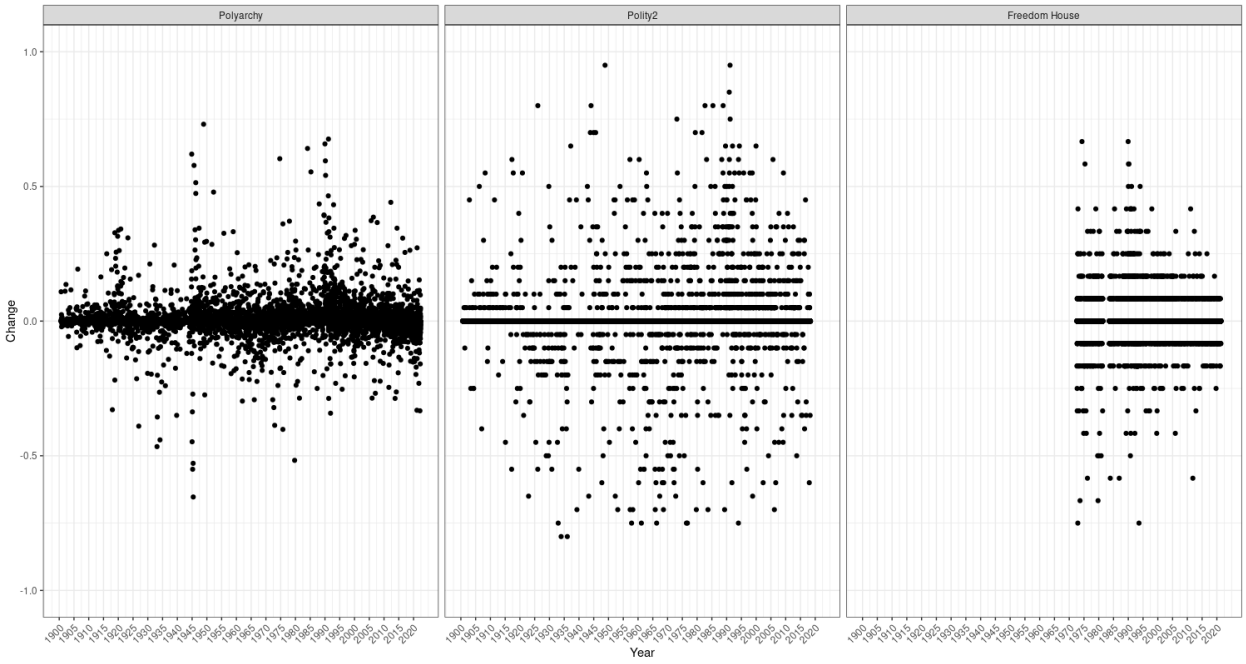


Table SI 10.3: Annual Change Statistics

	Polyarchy		Polity2		Freedom House
	1900-2022	1972-2022	1900-2018	1972-2018	1972-2022
Mean	0.003	0.004	0.003	0.006	0.002
Std. Dev.	0.046	0.045	0.082	0.079	0.067

SI 11 Freedom House predictions for different periods

Figure SI 11.1: Predicting 1988-2004

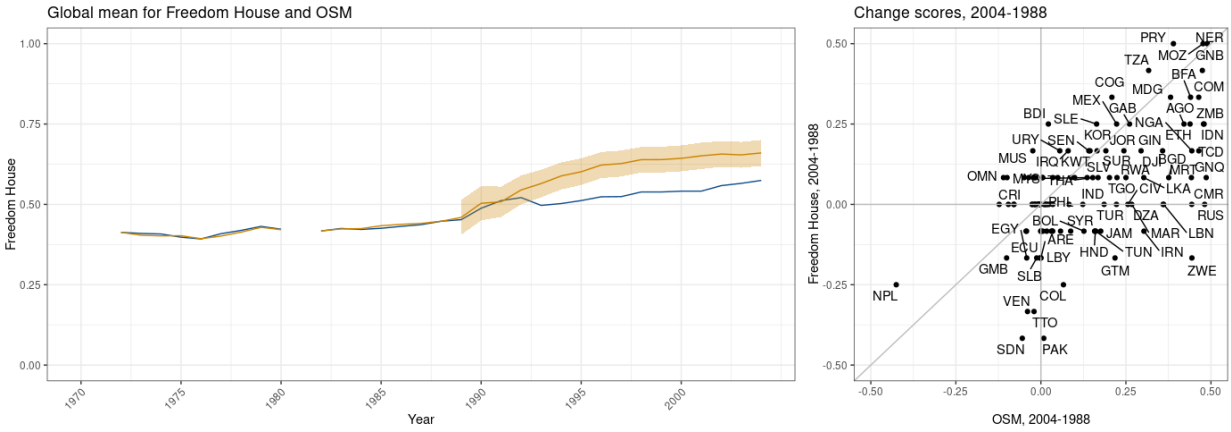
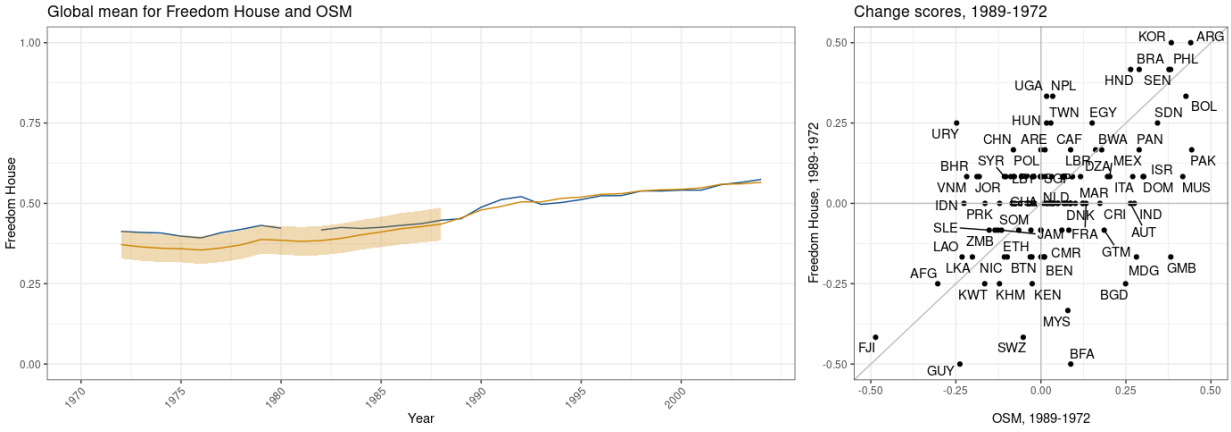


Figure SI 11.2: Predicting 1972-1988



SI 12 Predicting the interwar period

Figure SI 12.1: Polyarchy for the interwar period

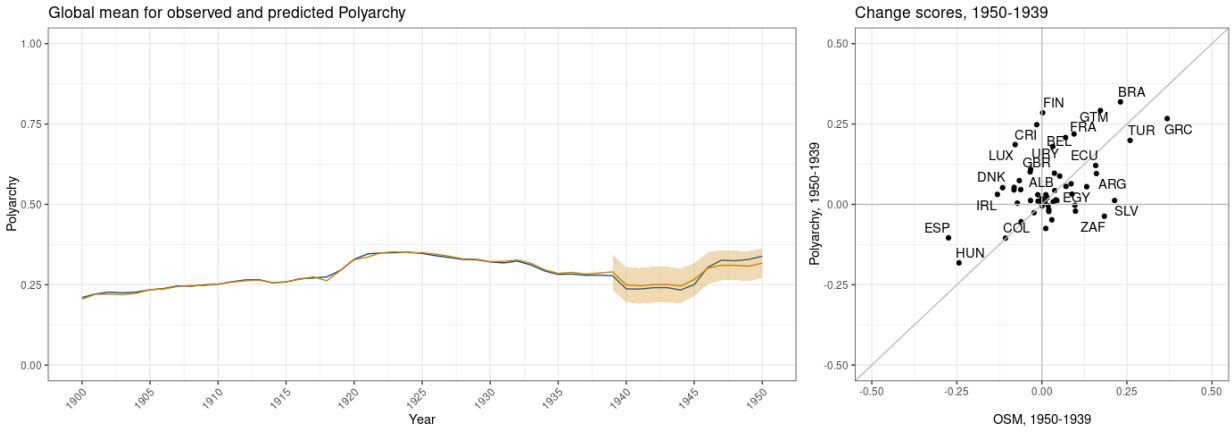
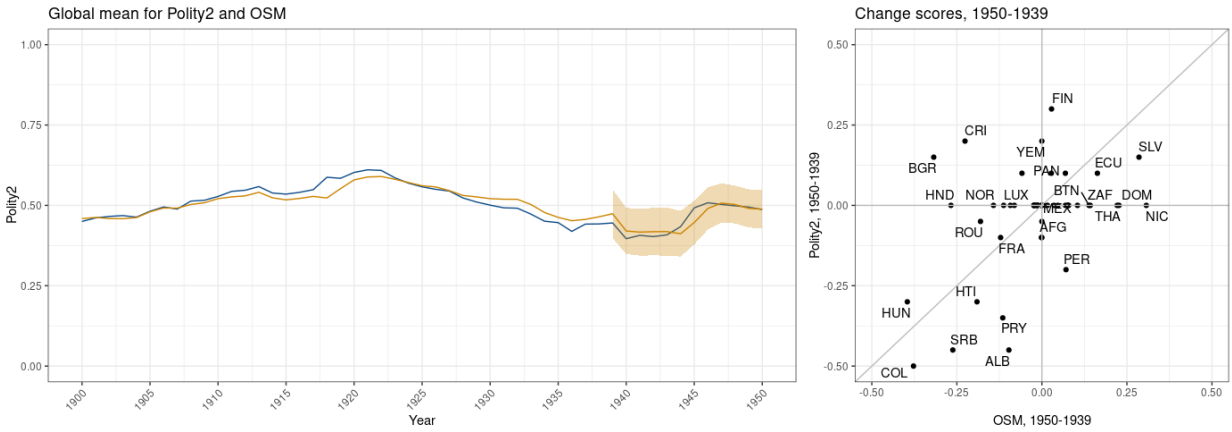


Figure SI 12.2: Polity2 for the interwar period



SI 13 Predicting the 1970s

Figure SI 13.1: Polyarchy for the 1970s

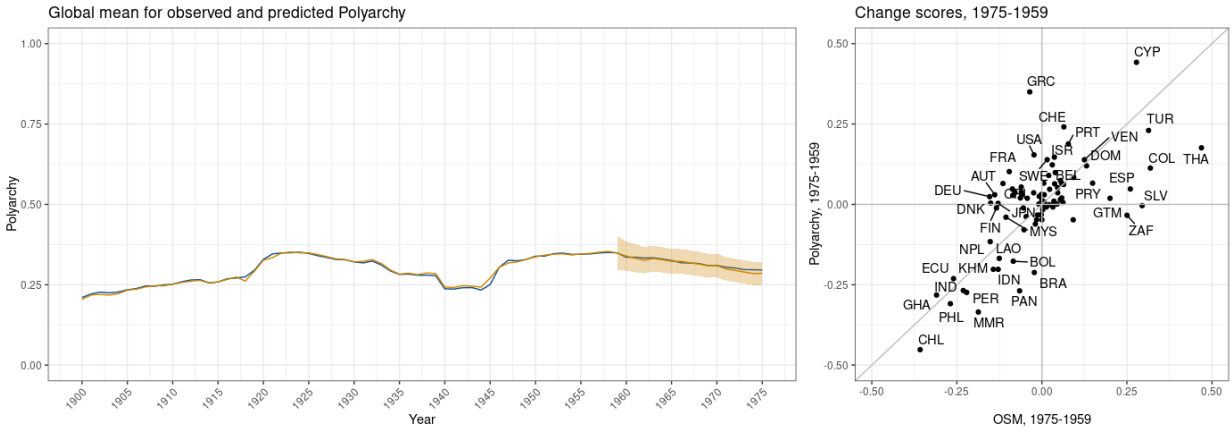
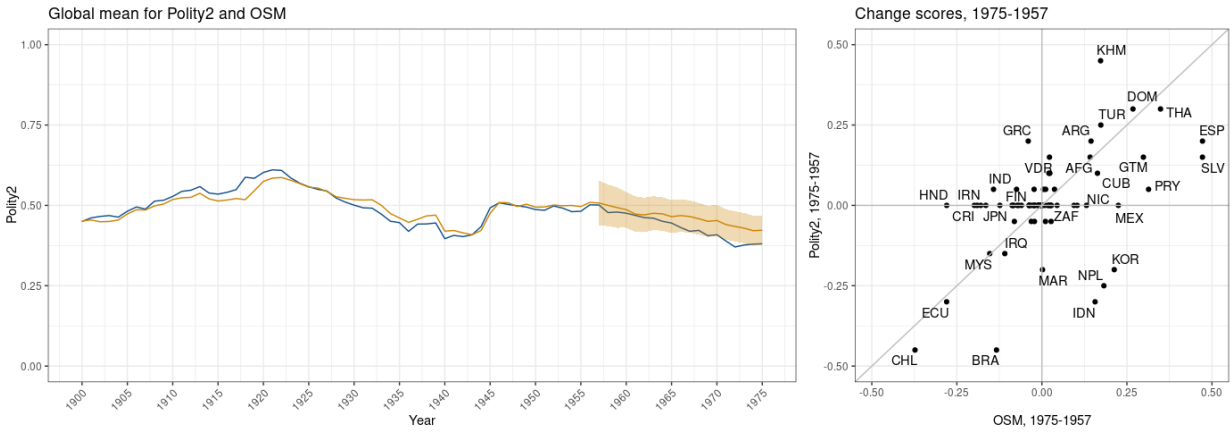


Figure SI 13.2: Polity2 for the 1970s



SI 14 Restricting the Polyarchy and Polity2 sample to 1972

Figure SI 14.1: Polyarchy since 1972

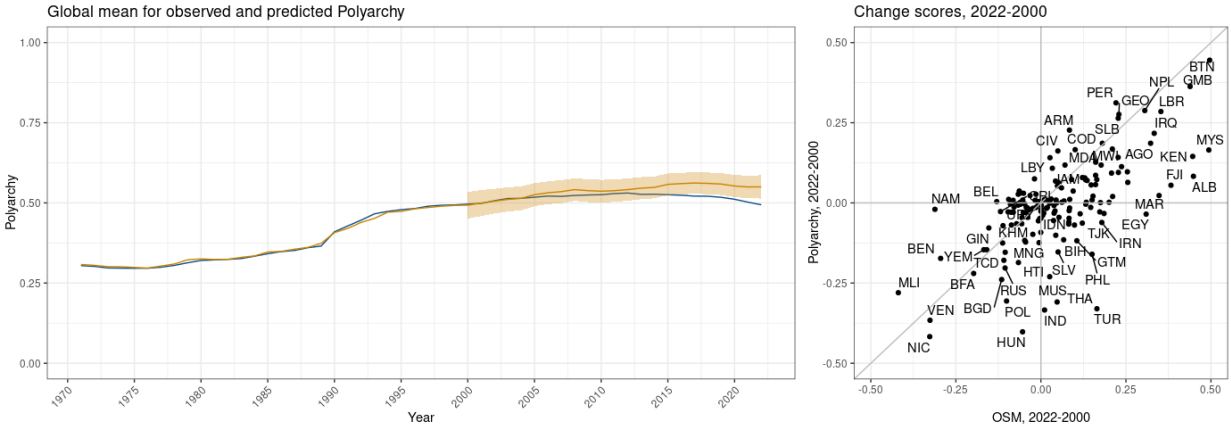


Figure SI 14.2: Polity2 since 1972