A Additional Empirical Analysis

A.1 Overview of commonly used data

Variable	Criteria
Freedom House	Political Rights: electoral process, political pluralism and participation, functioning of government. Civil Liberties: freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights.
Polity2	Competitiveness of executive recruitment, openness of executive recruitment, constraint on chief executive, competitiveness of political participation.
V-Dem: Electoral Democracy (v2x_polyarchy)	Universal suffrage, officials elected in free and fair elections, alternative sources of information and freedom of speech as well as freedom of association.
V-Dem: Liberal Democracy (v2x_libdem)	Criteria for electoral democracy plus constitutionally protected civil liberties, strong rule of law, independent judiciary, checks and balances that limit executive power.

Table 1: Summary of commonly used measures of democracy

A.2 Individual coder disagreement

In an impressive level of transparency, V-Dem releases data with anonymized individual-level coder decisions. In this section we explore trends in disagreement.

Figure 12 focuses on the coding choices of a single variable, the government censorship variable (v2mecenefm). The left panel plots the average number of coders per country for 1900-2022. This is typically over 5 since 1945, with a spike to around 11 in the 2010s. The middle panel plots the average confidence in coding decisions, which steadily increases in more recent years. Finally, the right panel plots the average standard deviation of the coding values for observations over the same time period. While this generally declines, the

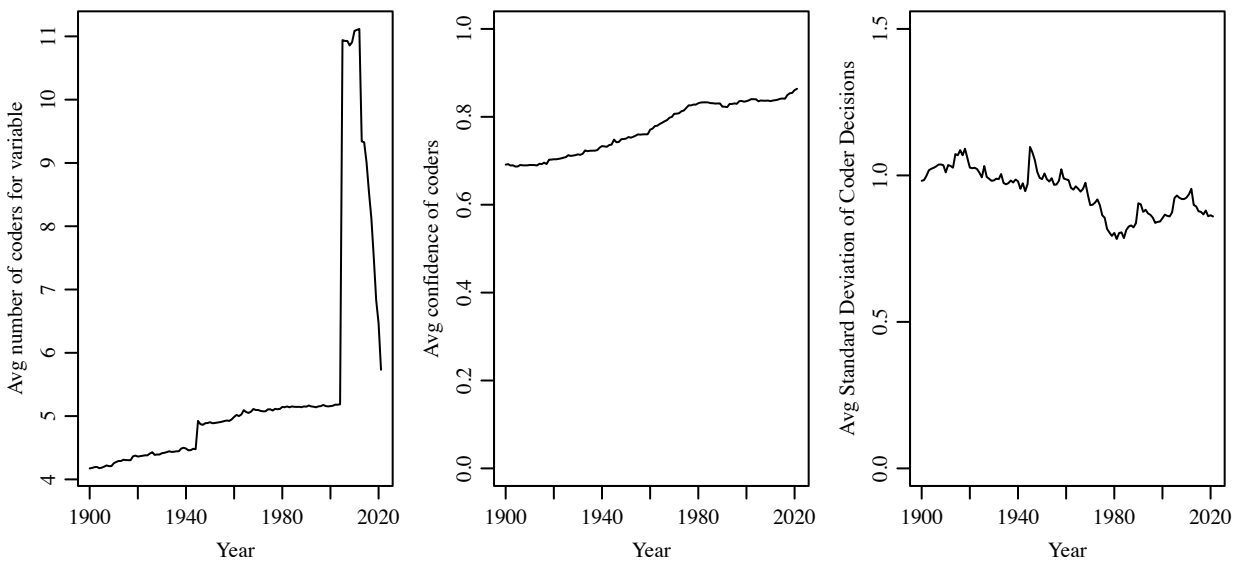

Figure 12: Individual coder data for government media censorship effort

trend is much more uneven than self-reported confidence, and disagreement on this variable has slightly increased from a low point in the 1980s. Perhaps more important, typical disagreement is substantial. This variable is coded on a 0 to 4 point scale, and the average standard deviation of coding decisions is a bit below 1. That is, on a typical observation for this variable, the typical code varies from the average decision by about 1 point on the scale.

Figure 13 shows the average standard deviation of coder decision by year for each variable which enters the three subindices of the Polyarchy score which rely on expert coding. Particularly those feeding into the freedom of expression and association (left and middle panels, respectively) tend to slowly decline, but the standard deviation is generally around 0.2. The variables in the clean elections index (right panel) tend to be more volatile since they are only coded for elections years but the average standard deviation is around the same level.

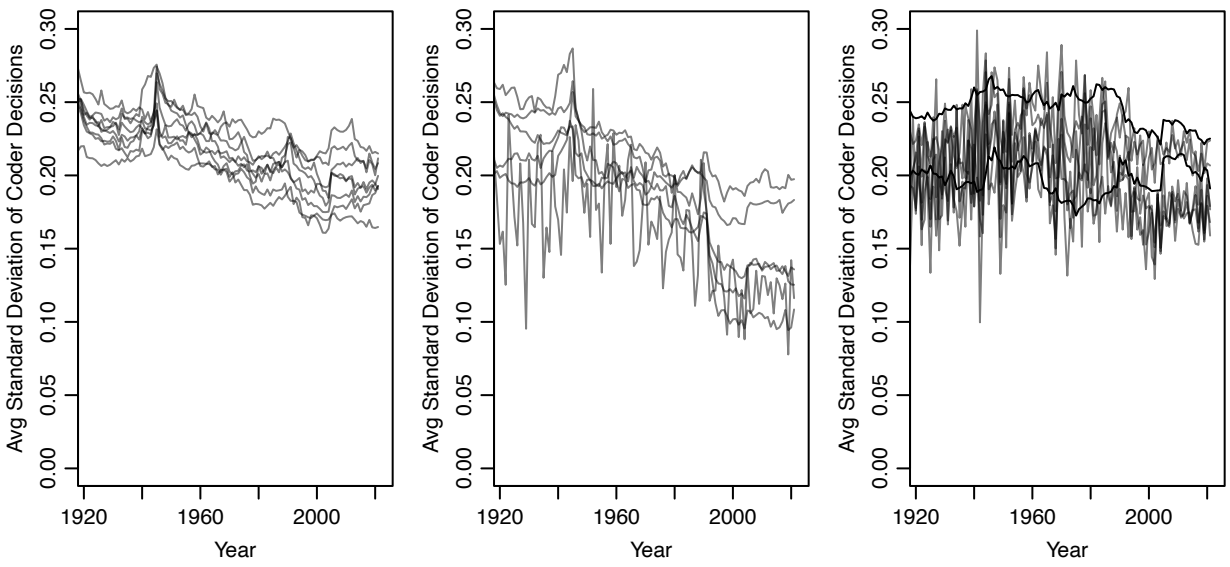

Figure 13: Standard deviation of yearly coder decisions for all variables which feed into the Freedom Of Expression (left), Freedom of Association (middle), and Clean Elections (right) indices.

A.3 Quantiles

The main text examines trends in the mean of some widely used democracy indices. To get a broader sense of the distribution, we can also look changes in the median and other percentiles.

Figure 14 displays the 10th to 90th decile of scores by year for the V-Dem polyarchy index (left) and Freedom House score (middle), and Polity (right) with the median (50th percentile) in a thicker line. For the polyarchy score, all deciles increased in the 1980s and 1990s, particularly in the middle range. Some of the deciles decline over the past decade, but generally do not fully reverse this trend. Like with the mean, the median trends up heavily from 1990 to the mid 2000s. Freedom house exhibits a similar pattern, albeit with less clear backsliding in recent years. For Polity, many of the deciles have continued to increase in the past decade.

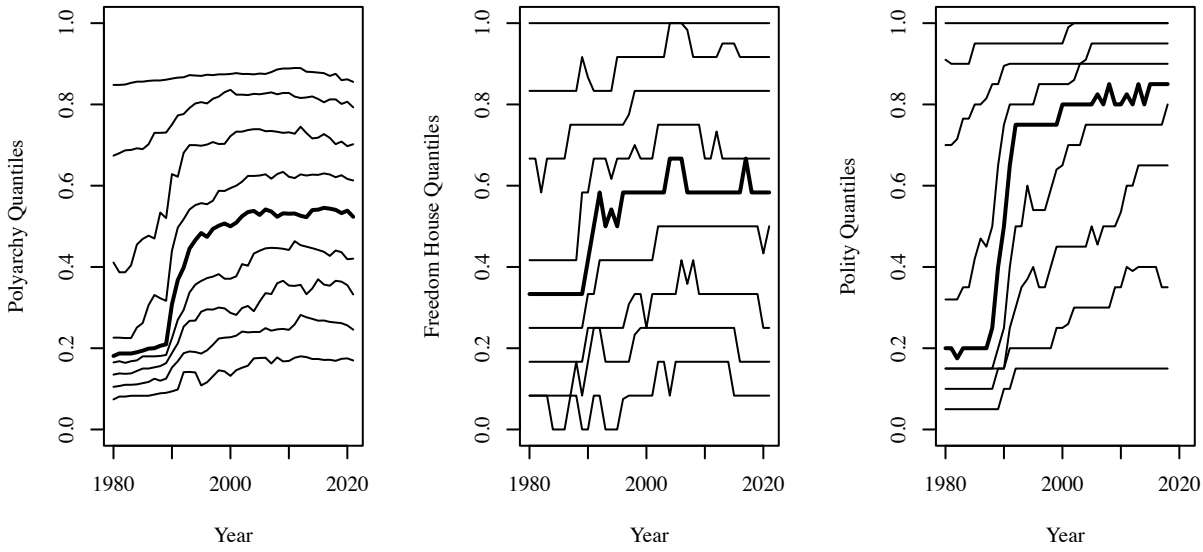

Figure 14: Percentiles of democracy scores

A.4 Comparison across V-Dem subindices

The widely used polyarchy index from V-Dem has five sub-indices: (1) freedom of expression, (2) freedom of association, (3) suffrage, (4) clean elections, and (5) elected officials. Three of these (the freedom of expression, freedom of association, and clean elections indices) are entirely comprised of expert-coded variables. The suffrage variable is objective, and the elected officials index (which broadly attempts to measure the share of political power held by elected individuals, setting aside any issues with those elections) is more objective than some of the other indices but requires some expert judgement about *de facto* procedures.

The left panel of Figure 15 plots the trend in the average of the three sub-indices which are relatively subjective. The freedom of expression and free association indices track each other quite closely, with a similar trend to the overall average. Both indicate some decline in the past decade, though to around the level in 2000, and still far above the level in 1990. The clean elections index is mostly increasing until the late 2010s, followed by a relatively

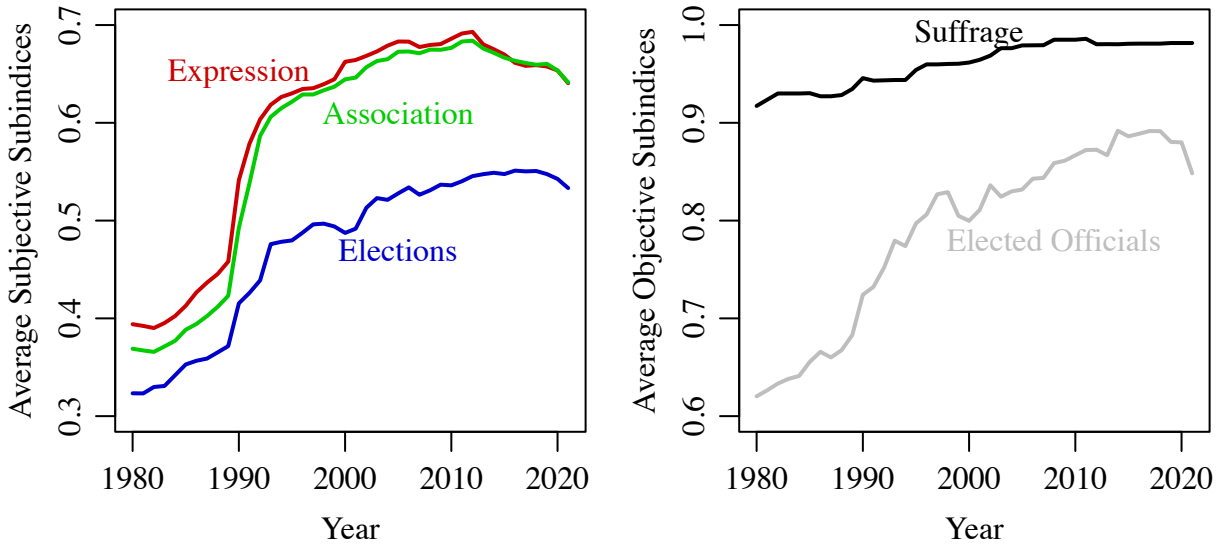

Figure 15: Trends in Subjective and Objective Sub-Indices of Polyarchy

small dip.[16]

More important is the contrast with objective indicators. As discussed above, suffrage is generally near universal in this time period, but it does increase steadily until 2000. The elected officials index increases quite a bit from 1980 to the mid 2010s, followed by a small dip in the around 2020.[17]

Overall, there are much less clear trends of backsliding on the more objective measures, as with meausures foced more on elections.

A.5 Election counts

Trends in the competitiveness of elections which are held could be misleading if incumbents are simply canceling or avoiding having elections in the first place. For example, if the most autocratic countries stopped holding elections, the average *held* election held would be

[16]It is also lower in absolute terms, but there is not too much meaning to different general levels across indices.

[17]It is helpful to note that this recent dip is due to a small increase in coups in the last few years, although in all cases coups have occurred only in very authoritarian countries such as Guinea, Mali, Afghanistan, Sudan, Myanmar, Chad and South Sudan.

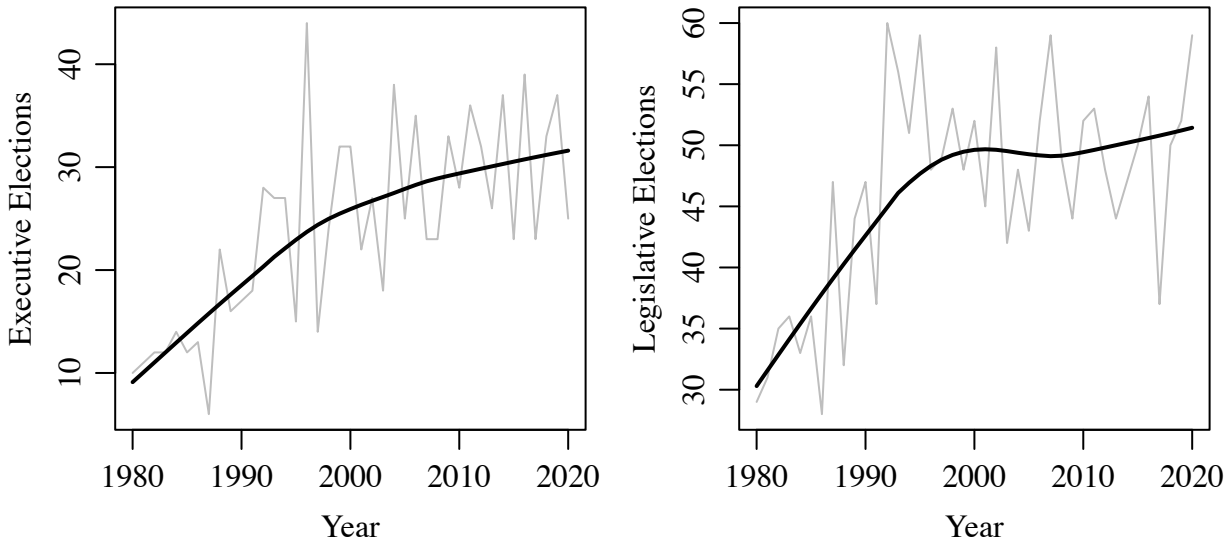

Figure 16: Count of executive and legislative elections

more democratic, even if the effect on global democracy is negative if not null.

Figure 16 show that this has not been occurring using NELDA data with two lowess smoothed trends. The left panel plots the count of executive elections by year, which generally increases in the early part. In the past decade or so the trend has continued if more slowly. The right panel plots the same for legislative elections. After a rapid increase up to 2000, the number of legislative elections has remained relatively constant.

Some of the early increase is driven by more countries entering the sample, but only three new countries are added after 2000.

A.6 Weighting of objective index

The main objective index gives equal weight to all components, some of which are arguably more important than others, and some of which are already a combination of several sub-indices. We start with the simple equal weight version for transparency. Figure 17 shows the robustness of our conclusion about general trends to different weightings. The thick black curve is the equal weight index, and the grey curves are the trend for 30 versions

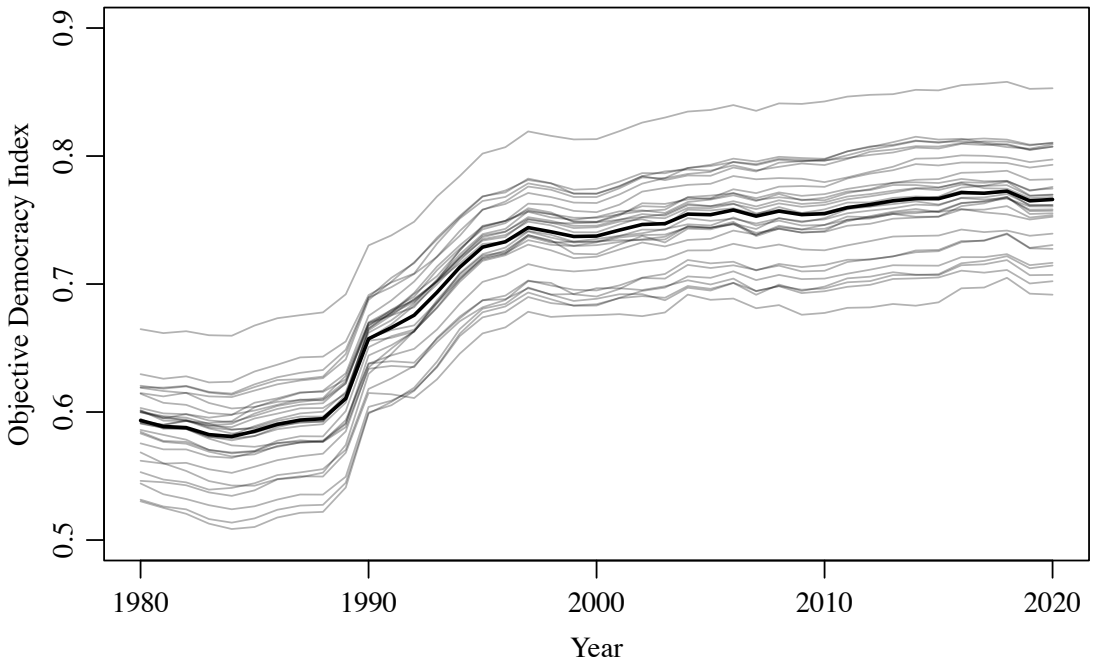

Figure 17: Trends in the objective index with random weights on the components.

of the index with randomly chosen weights on all components. In particular, weights are uniformly drawn on $(0, 1)$. While this moves the general index up and down somewhat, the general trends are similar.

A.7 Other indices by regime type

Figure 18 shows the average democracy score for Polyarchy (left panel) and Freedom House (right panel) subsetting by democracies and autocracies as classified by (Boix et al., 2013). In both cases, the average score within democracies is relatively stable. The increase in average scores during the third wave is driven by a combination of more countries becoming democratic, and autocracies becoming less undemocratic. The decline in average scores appears to be driven by by autocracies becoming even less democratic than democracies becoming less democratic.

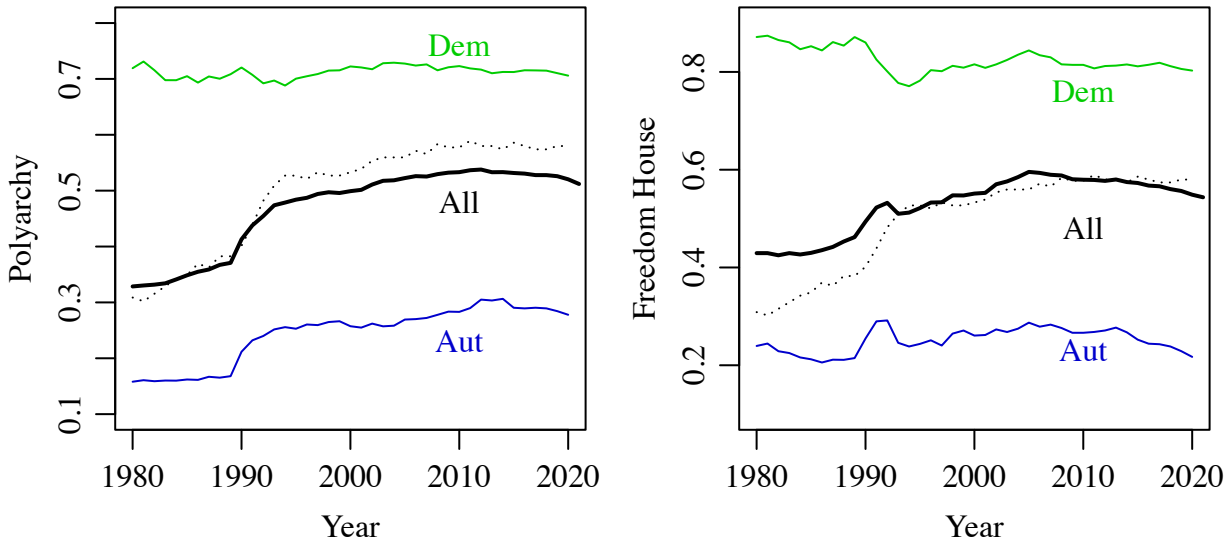

Figure 18: Trends in average Polyarchy score (left) and Freedom House democracy score (right) by regime type

A.8 Trends in decade changes

Figure 10 showed histograms of changes on several indices from the first to second half of decade ranges. To collapse this information, figure 19 plots the average and standard deviation of these changes for the objective index (left), Polyarchy (middle), and Freedom House (right) measures. For all three, the standard deviation of change is highest when the mean is highest during the third wave of democratization, meaning this is the period with the most volatility. On the objective index the standard deviation in the most recent decade is the lowest in the past 40 years. For the Polyarchy index and Freedom house there is some uptick in the standard deviation in recent periods, though though it is substantially lower than it was in the 1990s.

A.9 Regional Trends

We can also explore how the trends in different indices vary by region. Figure 20 shows the trends in the objective index, V-Dem polyarchy index, and the Freedom House democracy

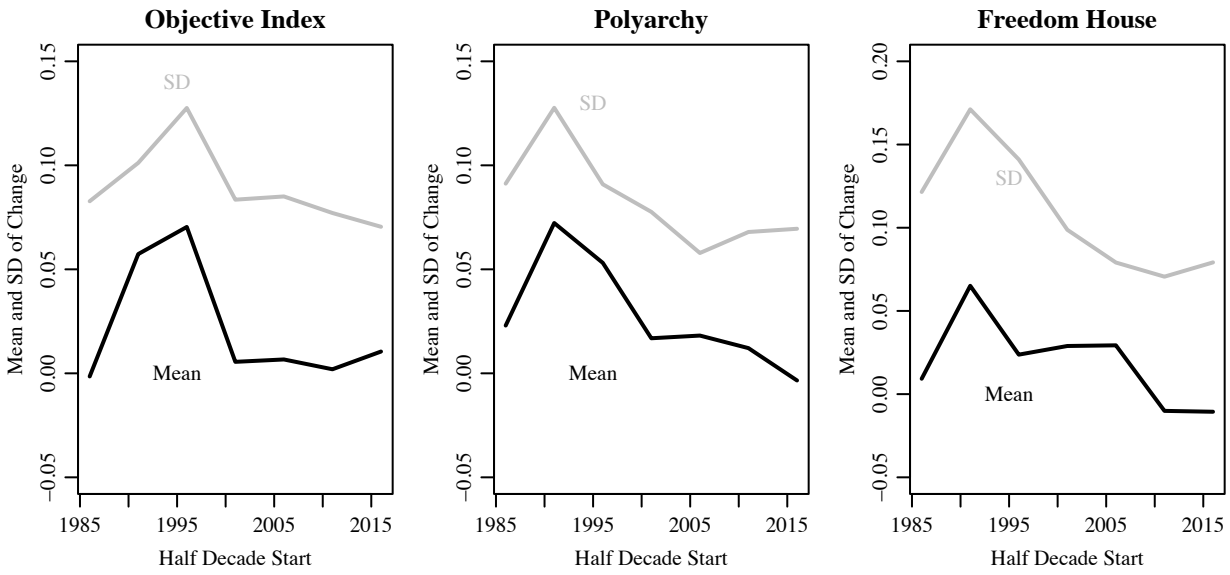

Figure 19: Trends in changes of indices from the first to second half of decade periods.

score for Western Europe and North America (orange), Latin America and the Caribbean (red), Asia and Pacific Islands (black), Eastern Europe and Central Asia (grey), Sub-Saharan Africa (blue) and Middle East/North Africa (green). Over the last 20 years the ordering of these regions is similar across the indices, and within each index there are few dramatic cross-regional differences. Eastern Europe and Central Asia is the only region that consistently declines over the past 20 years, while Asia and the Pacific Islands increased steadily throughout this time window until the late 2010s, when all three indices exhibited some decline. In all other regions, the objective index is generally flat if not somewhat increasing over the last two decades.

A.10 Balanced Panel

Particularly at the end of the Cold War, the set of countries in the world and hence in our data changes throughout our main time period. With the goal of summarizing democracy throughout the world it makes sense to include all countries in each year. However, it is also worth checking if trends are different for a balanced panel of countries that exist for our

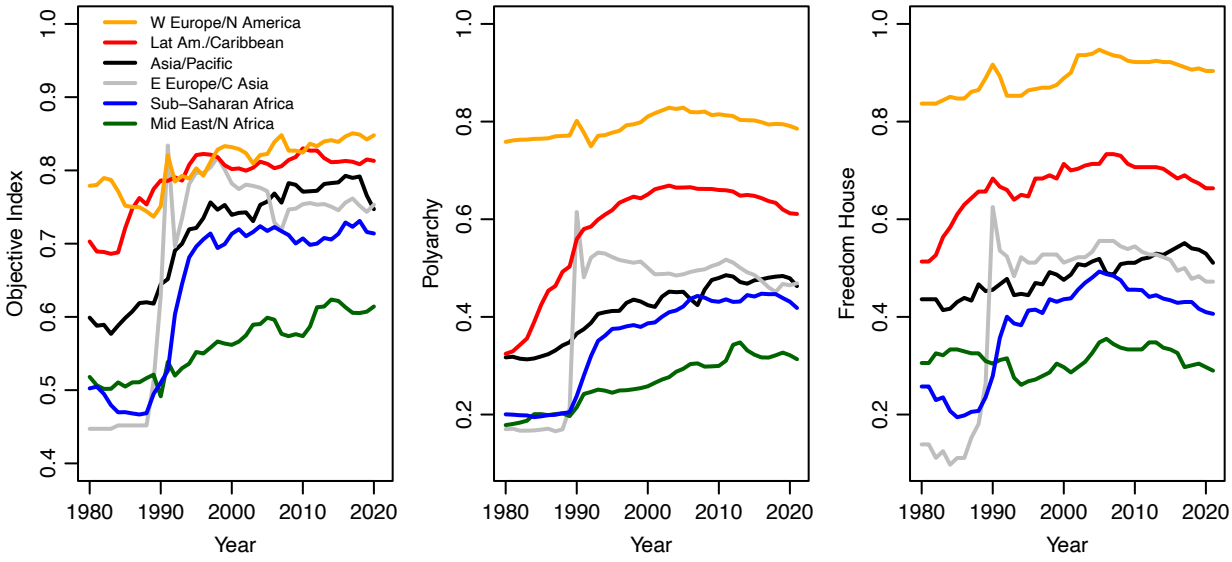

Figure 20: Trends in objective index, polyarchy index, and Freedom House democracy index by regions

entire time window.

Figure 21 reproduces two of our pain plots for this sample. The left panel shows the rate of incumbent party loss (in the previous election), and the right panel shows the objective index (thick line) as well as the V-Dem polyarchy index (thin black line) and Freedom House democracy score (thin grey line). The overall trends for this sample are quite similar.

B Additional variable descriptions

B.1 NELDA Data

For our main figure, we first create a country-year version of the NELDA data which pulls data from the last election held, *if* this election was held in the last six years.[18] We do three versions of this: including all elections, only elections where the incumbency was contested, and only elections where the incumbency was not at stake. This comes from the

[18]There are some cases where there is a long spell with no elections, for example Libya did not hold elections between 1965 and 2012. If we always pulled the last election, then the 1965 election would be included for over 40 years.

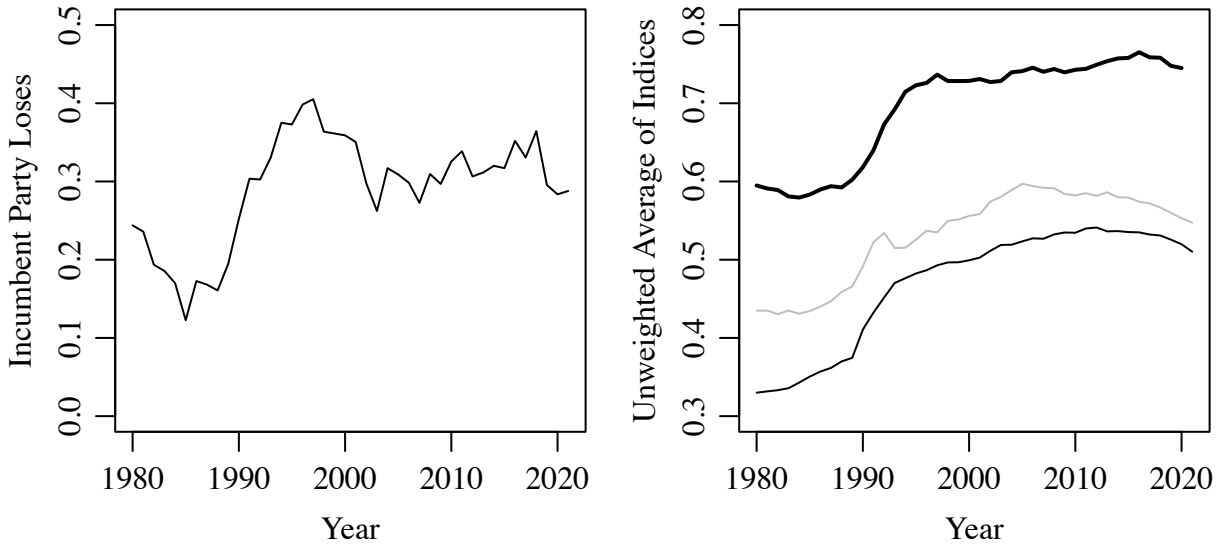

Figure 21: Rates of incumbent party loss (left) and the main indices (right) for a balanced panel.

Nelda20 variable ("Was the office of the incumbent leader contested in this election?"), and we count the three cases where this was coded as "N/A" or "Unclear" as not contested.

Our figures drop elections not held within the past 6 years, which removes about 6% of the country years for any election, about 11% of country years for elections where the incumbency was contested, and 18% of country years for elections where the incumbency was not contested. The patterns are very similar when including all country-years with a previous election or only including country-years where the previous election happened within the last 4 years.

The main variable we use for Figure 3 comes from Nelda24: "Did the incumbent's party lose?" We drop cases which are coded as "N/A" or "unclear", which generally correspond with instances where parties were banned or the leader does not have a party. In parliamentary systems, having to form a coalition does not constitute a loss for the leader's party.

B.2 DPI competitiveness index

Figure 4 includes a graph of the DPI competitiveness index for executive and legislative elections. This variable, called the Legislative and Executive Indices of Electoral Competitiveness (LIEC and EIEC), is described in the codebook as follows (pg. 14).

"To code the legislative index, countries are assigned the following scores:

1 No legislature

2 Unelected legislature

3 Elected, 1 candidate

4 1 party, multiple candidates

5 Multiple parties are legal but only one party won seats

6 Multiple parties DID win seats but largest party received more than 75% of the seats

7 Largest party got less than 75%"

The executive index is coded using the same scale (see pages 15-16 of DPI codebook for more details).

B.3 Executive constraints data

This section describes the executive constraints data, expanded from Meng (2020). The data includes the following three variables that document the presence of constitutional rules limiting executive power: (1) succession, (2) term limit, (3) dismissal. Each variable is a dummy variable, and takes a value of 1 if the constraint exists and 0 if the constraint does not exist.

The constitutional data was collected by a team of research assistants using the Comparative Constitutions Project portal. The portal includes copies of constitutions for all countries over all years after independence, and is kept up to date to include amendments, new constitutions, and notes periods where the constitution was suspended. We coded the

succession rule variable from scratch by reading constitutions for all countries and documenting when a constitution included a succession rule. The Comparative Constitutions Project Dataset (Elkins and Ginsburg, 2021) includes term limit and dismissal variables; we used CCP's variables as a base and filled in missing data.

Term limits Term limits constrain leaders by defining a time period after which they must step down or be ineligible for reelection. According to Svolik (2012, p. 198), "a term limit on a leader's tenure amounts to a line in the sand: Compliance is easily and publicly observable. Term limits thus embody a compromise about the limited authority of any single leader and provide an unambiguous signal of commitment to such a constraint."

Term limits most frequently designate a limited number of consecutive terms for which a leader may remain in office. For instance, Article 66 of the current constitution of Ghana includes the following clause: "A person shall not be elected to hold office as President of Ghana for more than two terms." Another common form of term limit is to prohibit the leader from remaining in office for consecutive terms while allowing for non-consecutive terms. Article 134.3 of the current constitution of Haiti states "The President of the Republic may not be re-elected. He may serve an additional term only after an interval of five years. He may in no case run for a third term." Since all of the executive constraints variables are dummy variables, We do not distinguish between different types of term limits in the main analysis, although we examine these differences in robustness checks.

In order for a term limit to be meaningful, the length of the term must be defined in the constitution. Term lengths generally range from 1 year to 8 years, although many constitutions do not define the length of the term. We require that the term length be defined as a finite number of years in order for the term limit variable to take a value of 1. In a small number of cases, the constitution has a term limit but does not define the length of the term; We recode the term limit variable to a 0 for those observations.

Succession A typical constitutional succession rule outlines formal procedures to replace the leader in the case of the incumbent's death, resignation, or permanent incapacity. Importantly, these rules identify an order of succession by designating an interim leader. For instance, the 1963 Kenyan constitution (Chapter II, Part I, Section 6) includes a clause that states:

"If the office of the President becomes vacant by reason of the death or resignation of the President... an election of a President shall be held within the period of ninety days immediately following the occurrence of the vacancy. While the office of the President is vacant, the functions of that office shall be exercised – by the Vice President."

Succession rules that clearly outline a designated successor can be thought of as a constraint on the executive by identifying an alternate center of power. Constitutional rules that lay out a succession hierarchy create a focal point by identifying the next leader, which solves the elite coordination problem surrounding the leadership transition. In fact, the very process of identifying the designated successor shifts the distribution of power in their favor by creating expectations that they are the heir (Meng, 2020). This so-called "Crown Prince problem" acts as a constraint on executive power.

Dismissal Another type of executive constraint are constitutions that include provisions for dismissing the leader. Leaders who can be dismissed on the basis of general dissatisfaction, crimes or misconduct, treason, or violations of the constitution are more constrained than those who do not face the possibility of dismissal. Rules allowing for dismissal are distinct from rules about leadership succession, and clauses that only mention dismissal via succession are not coded as a dismissal rule. For instance, Article 22L(3) of the current Singapore constitution states:

> "The Prime Minister or not less than one-quarter of the total number of Members of Parliament may give notice of a motion alleging that the President is perma-

nently incapable of discharging the functions of his office by reason of mental or physical infirmity or that the President has been guilty of –

(a) intentional violation of the Constitution;

(b) treason;

(c) misconduct or corruption involving the abuse of the powers of his office;

(d) any offence involving fraud, dishonesty or moral turpitude;

(e) intentionally or knowingly making a materially false or misleading statement of face, or intentionally or knowingly failing to state a material fact, to the Presidential Elections Committee for the purpose of demonstrating his eligibility to be elected as President"

C Formalization of theoretical explanations

In this section we formalize our explanations of what might be driving the difference between the expert-coded and more objective measures of democracy. We start with a purely statistical "coder model" which clarifies what kinds of bias would create a misleading trend in democracy scores. Next we present a strategic "leader model" to consider the possibility that more subtle democratic violations have become more common.

C.1 Coder model

Let θ_t be an indicator for how democratic a country is in year t. Multiple coders (indexed by j) code the country for multiple years (indexed by t).[19] Suppose we can write the score given by coder j to year t as:

$$s_{j,t} = \theta_t + \epsilon_{j,t}$$

[19]Note that we do not indicate the time at which the coding is done, which may affect the choice if standards change.

where $\epsilon_{j,t}$ is is a random variable with mean zero and variance σ_j^2. That is, the scores are unbiased, but noisy.

In this case, getting lots of expert opinions and averaging will give a very good estimate of the true democracy level. Assuming all expert's error terms follow the same distribution, taking the average of n expert scores generates a combined score which is has mean θ_t and variance σ^2/n. So, as n gets large, the average will be close to the truth.

Now suppose there is also a common shock to the expert scores η_t, and they can be written:

$$s_{j,t} = \theta_t + \eta_t + \epsilon_{j,t}$$

The common shock could reflect the fact that experts read common media sources, or publicly observed events occur which make a country seem more or less fundamentally democratic than it really is. Another possibility is that the outward appearance of how democratic a country is (even to experts) is misleading: perhaps a new leader comes to power who appears to be making positive democratic changes but behind closed doors is coopting opposition parties and buying off media outlets.

If we think of $\eta_t + \epsilon_{j,t}$ as the error term for expert j, these are now correlated and not necessarily mean zero because of the common component η_t. One implication of this is that even with a large number of experts, taking the average of the signals no longer approaches the truth because as $n \to \infty$ the average equals $\theta_t + \eta_t$. So, if η_t happens to be small or large (perhaps because all experts are relying on a common source of information they don't realize is biased), no number of experts can remove this component of the bias. Further, no measurement model can solve this problem since any two realizations of θ_t and η_t such that $\theta_t + \eta_t = \theta'_t + \eta'_t$ will produce the exact same vector of scores.

Change over time Given our interest in democratic backsliding, it is important to think about how this kind of common shock can affect inferences about changes in levels of democ-

racy.

Say we want to estimate the change in the level of democracy for the country between two years $t \in \{1, 2\}$. The natural estimate for expert j is:

$$s_{j,2} - s_{j,1} = \underbrace{\theta_2 - \theta_1}_{\text{real backsliding}} + \overbrace{\eta_2 - \eta_1}^{\text{change in bias}} + \underbrace{\epsilon_{j,2} - \epsilon_{j,1}}_{\text{change in noise}}$$

This is equal to what we want to know $(\theta_2 - \theta_1)$, with two distortion terms. Starting from the right, there is the change in noise, or the expert idiosyncratic terms. Under natural assumptions (e.g., the individual terms are mean zero, or at least have the same mean) this has mean zero, and if it is not correlated across experts, taking the average of their views will make the variance small. However, the change in bias term is common across experts, and so averaging their views (or using a more complicated measurement model) will not remove the influence of this term.

The key problem for detecting backsliding is not the existence of a common shock by itself, but the fact that the common shock may vary across time periods.[20] If so, democratic backsliding is not identified by this kind of estimator. A one unit increase in average expert assessment across time periods could mean a one unit increase in democracy with no change in bias, no change in democracy but an one unit increase in bias, or an infinite number of other combinations where the changes sum to 1.

C.2 Changing bias in the coder model

Here we formalize three related reasons why the bias term in expert perceptions may change across time: selective reporting, differing weights on different components of democracy, and motivated reasoning among coders.

[20] It is possible that objective measures are worse in the sense that the have a noisier error term, but are still better for detecting change if they do not carry the problem of time-varying common bias.

Open ended/selectively observed variables First, we consider the role of selective attention of the media and experts. This can lead to biased views of the state of democracy, even if all of the information they observe is verifiable and true, democracy indicator is clearly defined, and they are purely accuracy motivated. If some coding decisions are made about more recent times and others for the distant past, related issues of what facts are *remembered* may arise.

Again, consider how democratic a country is in time t, θ_t, and in this section assume θ_t is between 0 and 1. An expert observes a large number of facts about the country θ_t^k ("did the loser of the election accept defeat"). Assume more democratic countries produce a higher rate of "democratic" facts, where $\theta_t^k = 1$ means fact k indicates the country is democratic. To simplify, suppose the level of democracy is the probability that an individual fact is democratic: $Pr(\theta_t^k = 1) = \theta_t$.

If observing a large number of facts, a natural estimate for θ_t is just the share of democratic facts.[21] E.g., if n_t facts are observed, and $n_{t,1}$ are 1s, then the estimate of how democratic the country is will be $n_{t,1}/n_t$. By standard results, in large samples this estimate will converge to the truth.

Now suppose that some types of facts are more observable than others. In particular, suppose the typical democratic fact is observed with probability $p_{1,t}$ in time t and the typical undemocratic fact is observed with probability $p_{0,t}$.[22] For example, if acts of democratic backsliding tend to be more discreet Luo and Przeworski (2019), it may be the case that $p_{0,t} < p_{1,t}$.

[21] Under various assumptions (e.g., the prior follows a uniform distribution), the mean of the estimate about θ_t after observer a large number of observations is in fact the mean of the Bayesian posterior belief about θ_t.

[22] We could also add j subscripts/superscripts if there are multiple experts.

Then the proportion of democratic facts conditional on being observed is:

$$\frac{\theta_t p_{1,t}}{\theta_t p_{1,t} + (1-\theta_t) p_{0,t}}$$

If $p_{1,t} = p_{0,t}$, this reduces to θ_t. So, imperfect observation doesn't necessarily lead to biased inferences unless some kinds of facts are more observable than others.

If $p_{1,t} > p_{0,t}$, then the observed facts will be skew towards democratic ones, and if $p_{1,t} < p_{0,t}$ then the opposite will be true. A sophisticated observer who knows this could still back out the true value of θ_t.[23] However, there is substantial evidence that even highly educated subjects often do not always properly adjust for selection problems in the samples they observe (see Enke 2020 for a prominent recent example, and Brundage et al. 2022 for an overview). If the observer is subject to this *selection neglect* and takes the observed facts as representative, their opinions will be biased away from the truth when $p_{1,t} \neq p_{0,t}$.[24]

This could lead to incorrect views of backsliding if the observability parameters change (e.g., democratization or backsliding stories start to get more traction), as this will change the distribution of observed events even if the underlying democracy variables are constant. More concretely, there are at least two plausible mechanisms through which this can happen:

1. If improved media technology makes it easier to observe previously hidden undemocratic facts and actions, this can lead to the perception of countries getting less democratic. For example, suppose $p_{1,t} = 1$ or is constant; plausible as leaders always aim to

[23] In particular, if they observe a large number of facts and share q are democratic, then then can solve:

$$q = \frac{\theta p_1}{\theta p_1 + (1-\theta) p_0}$$
$$\theta = \frac{q p_0}{q p_0 + (1-q) p_1}$$

[24] One insight from this is that such a bias is largest when θ_t is intermediate. As long as $p_{1,t} > 0$ and $p_{0,t} > 0$, as $\theta_t \to 0$ the proportion of undemocratic observed facts goes to 0 (the truth) and when $\theta_t \to 1$ it approaches 1 (again, the truth). As a result, this bias will tend to be acute when (1) there is a bias in what gets covered, and (2) for middling democracies.

broadcast their democratic bona fides. However, suppose that with improved technology, it is easier to detect and spread information about undemocratic action, and so $p_{0,t}$ is increasing in t. Then, even if θ_t is constant, the proportion of observed democratic facts will decline.

2. When the dominant "narrative" in the media is about democratization we may see more stories about that, and when the dominant narrative is about backsliding there will be more stories about that. This can create a feedback loop where some real democratization/backsliding can change how other countries are perceived by changing news coverage about them.

Subjective definitions/weights Another way to think about how selective reporting and attention can lead to time vary-ing bias is that some variables are hard to define, or combine different components. To keep things simple and highlight this particular problem (even though it typically coexists with others), we assume there is no noise in observation here; i.e., all experts have the same "hard data" to work with.

Suppose an indicator θ_t has two components θ_t^1 and θ_t^2. For example, if the indicator is whether parties are banned from running, θ_t^1 could mean whether the largest opposition party is banned while θ_t^2 corresponds to a more minor party. Banning party 1 might be more important, but it is hard to know how much this should be given more weight (if at all).

Formally, let $w_{j,t}$ be the weight that coder j places on component θ_t^j in year t. The score they report will be:

$$\theta_{j,t} = w_{j,t}\theta_t^1 + (1 - w_{j,t})\theta_t^2$$

I.e., we are assuming the weights sum to 1, which reduces the number of variables to keep track of. Clearly if $\theta_t^1 \neq \theta_t^2$ then coders with different weights will disagree even if they agree on the individual components.

Further, changes in democracy scores could be purely determined by changes in weights. To simplify suppose the indicators remain constant across years 1 and 2 (and so drop the t subscripts on these), but the weights potentially change. Then:

$$\theta j, 2 - \theta j, 1 = (w_{j,2} - w_{j,1})\theta^1 + (w_{j,1} - w_{j,2})\theta^2 = (w_{j,2} - w_{j,1})(\theta^1 - \theta^2)$$

If the country is equally democratic on both components ($\theta^1 = \theta^2$), the change will be zero regardless of the weights. However, if not, a changing weight can lead to a change in the democracy weighting equal to the product of the weight change and the difference on the components.

Potential sources of changing weights could relate to the previously discussed biases. If the media focuses more on particular components of democracy (perhaps because they are dimensions with salient changes), that can increase the perceived importance of that component. For example, if a president attacks the independence of the judiciary, that will likely get attention in the media; and with good reason. However, this increased attention could lead coders to pay more attention to judicial independence when thinking about executive constraints or democracy more generally. Such a bias will magnify any real changes in the underlying indicators.

It is worth noting that one potential way to reduce this bias is to ask more specific questions, which is a main benefit of the V-dem data over other expert-coded measures.[25] Still, we think future data collection efforts can go further on this front and ask even more specific and ideally objective questions.

Motivated Beliefs Finally, we consider how how coders wanting to believe certain things mre generally can skew their beliefs and bias estimates of backsliding based on changes in their reports.

[25] See https://www.v-dem.net/about/faq/.

To do so we need to add a bit more structure to the expert beliefs. Suppose each expert has an objective belief which is normal with mean s_j and variance σ_j^2. Using the formalization of motivated beliefs from Little (2021), let f_j be the density of this objective belief, and suppose the subjective (motivated) belief is given by:

$$\tilde{f}_j = \arg\max_{f' \in \mathcal{F}} -D_{\mathrm{KL}}(f' \| f_j) + \int v(\theta) f'(\theta) d\theta$$

where $\mathcal{F}$ is the set of all density functions on $\mathbb{R}$, $kl(f' \| f_j)$ is the Kullback-Leibler divergence from the objective belief to the subjective belief. This generates a penalty for moving from the objective belief, capturing an accuracy motive for the expert; see Little (2021) for an extensive discussion of why this is a good measure for the accuracy motive. Finally, $v(\theta)$ is a directional motive indicating how much the expert likes believing the democracy level of the country is θ.

Theorem 4 in Little (2021) states that as long as $\int f_j(\theta) e^{v(\theta)} d\theta$ is finite—intuitively, this will hold unless the expert likes holding extreme beliefs—the motivated belief is given by:

$$\tilde{f}_j(\theta) = \frac{f_j(\theta) e^{v(\theta)} d\theta}{\int f_j(\theta') e^{v(\theta')} d\theta'}$$

A simple v function to work with is a linear one, where $v(\theta) = d_j \theta$, where $d_j > 0$ indicates wanting to think the country is more democratic and $d_j < 0$ means wanting to think the country is less democratic. If so the motivated belief becomes normal with mean:

$$s_j + d_j \sigma_j^2$$

and variance σ_j^2. That is, relative to the objective belief, the motivated belief is shifted up or down by a distortion equal to $d_j \sigma_j^2$. Unsurprisingly, this magnitude of the distortion is increasing in $|d|$.

Less obvious, it is also increasing in in σ_j^2, meaning that the motivated beliefs matter more when there is more uncertainty about the true value. This is arguably true for "more subjective" variables. If so, the more subjective a variable, the more we need to worry that motivated beliefs can cause more biased assessments and excessive disagreement among experts.

It is also feasible that directional motives change across time: perhaps an expert dislikes the new government that takes office and wants to think their behavior is undemocratic, or the expert perceives a general regional or global trend and wants to think that the country they are assessing is a part of that. If so, again adding t subscripts,[26] the motivated belief in time t for expert j has mean:

$$s_{j,t} + d_{j,t}\sigma_j^2$$

The change in the belief across years is then:

$$s_{j,2} - s_{j,1} = \underbrace{\theta_2 - \theta_1}_{\text{real backsliding}} + \overbrace{(d_{j,2} - d_{j,1})\sigma_j^2}^{\text{change in directional bias}} + \underbrace{\epsilon_{j,2} - \epsilon_{j,1}}_{\text{change in noise}}$$

Again, averaging across experts can remove the bias introduced by a change in the idiosyncratic terms. However, if experts directional motives systematically change over time, this can lead to the appearance of a change in how democratic a country is absent any real change. A related possibility is that the directional motives are constant, but the variance of the belief decreases over time as coders have more information. For any $d_j \neq 0$, this will create a time-varying bias as well since the distortion will get smaller if the variance is decreasing in t.

[26]For simplicity we are assuming the beliefs about the democracy levels across time are uncorrelated, which is clearly unrealistic, but the core insights would still hold, need to prove this.

C.3 Leader Model

Undemocratic behavior is a strategic choice made by politicians, and whether their behavior is reported may influence the choices they make. Further, a possible objection to the analysis so far is that it does not capture the idea that the nature of democratic erosion may be changing in ways that are hard to detect with objective measures. We now present a decision-theoretic model to explore this idea.

Consider a leader who is deciding how much undemocratic behavior to engage in. Treat this as a continuous choice with two dimensions: subtle undemocratic action $a_s \geq 0$ and blatant undemocratic action $a_b \geq 0$. Let $a = a_s + a_b$ be the sum of the actions, and assume the return can be written as a function of this sum: $v(a)$. That is, we assume these kinds of actions are interchangeable on the benefit side.

Undemocratic actions also carry two kinds of costs, both of which may be different for blatant and subtle choices. The first are direct costs which apply whether or not the action is reported (e.g., think of this as internal pressure from antagonizing domestic constituencies, salaries paid to agents who manipulate elections, etc.). The second costs are only applied when the undemocratic behavior is reported (this can also be interpreted as undemocratic behavior being "detected").

Suppose the direct costs to blatant and subtle antidemocratic actions are $k_b(a_b)$ and $k_s(a_s)$, respectively. Allowing these functions to differ is important given we assumed they are interchangeable on the benefit side. To be concrete, suppose $v(a) = \alpha + \beta a$ is the expected incumbent vote share with total anti-democratic action a. Then we can interpret $k_b(a)$ as the direct cost of increasing the expected vote share by βa units using blatant means, and $k_s(a)$ as the cost of achieving the same benefit through subtle means. That is, despite the way the value function is written, we are not assuming these actions are equally effective.

We assume there is an additional cost to *reported* undemocratic action. Let the cost of reported action type $j \in \{s, b\}$ be $rp_j a_j$, for $r \geq 0$ and $p_j \geq 0$. Think of p_j as the

rate of reporting this kind of action, which could correspond to the extensive margin (the probability that an action is reported) or the intensive margin (how much media coverage and other attention is paid to the violation). The $r > 0$ parameter captures the cost of reported violations. To formalize the distinction between subtle and blatant actions, assume the latter are reported more: $p_s \leq p_b$.

Writing out the full utility function:

$$u_L(a_s, a_b) = v(a_s + a_b) - k_s(a_s) - k_b(a_b) - r(p_s a_s - p_b a_b)$$

Assume that v is continuous, strictly increasing and weakly concave in a. This implies there are diminishing returns to undemocratic action, and also implies that the two actions are substitutes in the sense that doing more of one type of action decreases the return to the other. Assume the direct costs k_s and k_b are continuous, with increasing marginal cost ($k_j'' > 0$). To ensure an interior solution we also assume the marginal cost is unbounded ($\lim_{a_j \to \infty} k_j'(a_j) = \infty$).

What is to be explained? To contrast with the possibility raised in the previous model – that the decline in expert-coded democracy scores is just driven by coder bias – here we consider what changes in exogenous parameters might increase subtle democratic violations which are not detected by objective measures.

In particular, some questions we can ask which can be mapped to the observed trends are: How does changing different exogenous parameters change (1) the solution to the optimal choices (a_s^*, a_b^*), (2) reported violations $(p_s a_s^*, p_b a_b^*)$, and (3) the aggregate amount of antidemocratic action $(a^* = a_s^* + a_b^*)$ or returns to antidemocratic action $(v(a^*))$?

In particular, we may think that objective measures may capture blatant violations (a_b^*). Subjective measures may do a better job of capturing subtle violations (a_s^*) if expert coders have a strong sense that democratic violations are happening behind closed doors or in ways

that objective metrics may not capture. However, subjective measures may also be more sensitive to the amount of reporting (of both blatant and subtle violations). Finally, the "outcome" variables like turnover and incumbent vote shares plausibly reflect the return to violations $v(a^*)$.

Solution The first-order conditions for the two choices are:

$$v'(a_s + a_b) - k_s'(a_s) - rp_s = 0 \tag{1}$$

$$v'(a_s + a_b) - k_b'(a_b) - rp_b = 0 \tag{2}$$

Given the assumptions placed on these functions, a unique solution exists to this system of equations.[27] To simplify we focus on the case where the solution is interior; i.e., the leader picks a strictly positive about of subtle and blatant violations ($a_j^* > 0$).[28]

To see how the optimal choices change as the exogenous parameters change, we need to implicitly differentiate equations (1)-(2). We consider three possible changes.

More pressure First, we consider an increase in the cost to reported violations r. This could represent increased pressure from foreign countries, businesses, or donors to rule in a more democratic fashion.

Proposition 1. *As the cost of reported violations (r) increases:*

i blatant undemocratic actions a_d^* and total undemocratic actions a^* decrease,

ii subtle undemocratic actions increase if and only if p_s is sufficiently small and decrease otherwise

[27] In particular, the objective function is globally concave in a_s and a_b with a positive semidefinite Hessian, which ensures that any solution to the FOC is a unique global maximizer.

[28] We can ensure that at least one action is taken by assuming that $v'(0)$ is sufficiently large. The solution may only involve one type of action if one is much cheaper than the other; capturing this case just requires adding some caveats to the main results.

Proof Totally differentiating the FOC (equations 1 and 2) with respect to r gives:

$$v''(a)\frac{\partial a_s}{\partial r} + v''(a)\frac{\partial a_b}{\partial r} - k_s''(a_s)\frac{\partial a_s}{\partial r} - p_s = 0 \tag{3}$$

$$v''(a)\frac{\partial a_b}{\partial r} + v''(a)\frac{\partial a_s}{\partial r} - k_b''(a_b)\frac{\partial a_b}{\partial r} - p_b = 0 \tag{4}$$

Solving for $\frac{\partial a_b}{\partial r}$ and $\frac{\partial a_s}{\partial r}$ gives:

$$\frac{\partial a_b^*}{\partial r} = \frac{-p_b k_b''(a_b^*) + v''(a^*)(p_s - p_b)}{k_s''(a_s^*)k_b''(a_b^*) - k_s''(a_s^*)v''(a^*) - k_b''(a_b^*)v''(a^*)}$$

$$\frac{\partial a_s^*}{\partial r} = \frac{-p_s k_s''(a_s^*) + v''(a^*)(p_b - p_s)}{k_s''(a_s^*)k_b''(a_b^*) - k_s''(a_s^*)v''(a^*) - k_b''(a_b^*)v''(a^*)}$$

Both derivatives have the same denominator, which is positive since all three individual terms are positive. So the sign of both derivatives is equal to the sign of the numerator. For the change in blatant undemocratic behavior, both terms in the numerator are negative (since $k_s''(a_s) > 0$, $v''(a) < 0$, and $p_s < p_b$), and so increasing r leads to less of this type of violation.

For subtle undemocratic behavior, the first term in the numerator is negative. The second term is positive however, reflecting the fact that making caught violations more costly also deters blatant violations, and since these actions are substitutes this increases the marginal return to subtle violations. In the extreme, as $p_s \to 0$ (subtle violations are never caught), increasing the penalty for caught violations will increase subtle anti-democratic action because of the substitution effect. However, if p_s is large than the direct effect dominates and subtle violations will decrease, completing part ii.

To complete part i, the change in a^* as r increases is:

$$\frac{\partial a^*}{\partial r} = \frac{\partial a_s^*}{\partial r} + \frac{\partial a_b^*}{\partial r} = \frac{-(p_b + p_s)k_b''(a_b^*)}{k_s''(a_s^*)k_b''(a_b^*) - k_s''(a_s^*)v''(a^*) - k_b''(a_b^*)v''(a^*)} < 0$$

■

Intuitively, increasing r makes both kinds of violations more costly, but also makes subtle violations relatively less costly. So, there is a decrease in aggregate undemocratic behavior, but also some substitution to subtle actions. This unambiguously decreases blatant undemocratic behavior but can increase or decrease subtle undemocratic behavior.

Even if increasing r increases subtle democratic violations, and we accept that expert coders can detect subtle violations while objective measures cannot, this alone is not a compelling explanation for the observed patterns since aggregate violations should decrease. This is inconsistent with both expert coders becoming more pessimistic about on average and the key empirical fact that incumbent parties are not winning elections at a higher rate or with higher vote shares.

More Reporting As discussed in the coder model, another potentially realistic change over the past 10 years is improved information and communication technology may render democratic violations more visible. A natural way to model this is increasing the p parameters, particularly p_s (as blatant violations are likely easy to detect in general). Focusing just on subtle violations:

Proposition 2. *As the reporting subtle violations p_s increases:*

i Subtle undemocratic actions (a_s^*) and total undemocratic actions (a^*) decrease,

ii Blatant undemocratic actions and reported blatant undemocratic actions (a_b^* and $p_b a_b^*$) increase

iii Reported subtle undemocratic actions $p_s a_s^*$ can increase or decrease

Proof To examine the effect of increasing p_s, totally differentiating the FOC with respect

to this parameter gives:

$$v''(a)\frac{\partial a_s}{\partial p_s} + v''(a)\frac{\partial a_b}{\partial p_s} - k_s''(a_s)\frac{\partial a_s}{\partial p_s} - d = 0 \tag{5}$$

$$v''(a)\frac{\partial a_b}{\partial p_s} + v''(a)\frac{\partial a_s}{\partial p_s} - k_b''(a_b)\frac{\partial a_b}{\partial p_s} = 0 \tag{6}$$

Solving for $\frac{\partial a_b}{\partial p_s}$ and $\frac{\partial a_b}{\partial p_s}$ gives:

$$\frac{\partial a_s^*}{\partial p_s} = \frac{r(v''(a^*) - k_b''(a_b^*))}{k_s''(a_s^*)k_b''(a_b^*) - k_s''(a_s^*)v''(a^*) - k_b''(a_b^*)v''(a^*)} < 0$$

$$\frac{\partial a_b^*}{\partial p_s} = \frac{-rv''(a^*)}{k_s''(a_s^*)k_b''(a_b^*) - k_s''(a_s^*)v''(a^*) - k_b''(a_b^*)v''(a^*)} > 0$$

Adding these two gives $\frac{\partial a^*}{\partial p_s} < 0$ ■

The intuition here is that raising the observability of subtle undemocratic action raises this cost of using this means, leading to some substitution to blatant action (but an overall decrease). For *reported* actions, there are countervailing effects where increasing p_s leads to more reporting of each subtle action, but decreases the number of these actions. So it is possible that more subtle undemocratic actions are observed even if less are occurring.

Subtle subversion becomes easier A final change we might want to capture is if subtle subversion becomes easier. One way to formalize this is to scale down the direct cost of subtle undemocratic actions, $k_s(a_s)$:

Proposition 3. *As subtle undemocratic actions become easier:*

i Subtle undemocratic actions, (a_s^*) reported subtle undemocratic actions ($p_s a_s^*$) and total undemocratic actions (a^*) increase,

ii Blatant undemocratic actions and reported blatant undemocratic actions (a_b^* and $p_b a_b^*$) decrease

Proof To study the effect of subtle behavior becoming easier, write the utility function as

$$u_L(a_s, a_b) = v(a_s + a_b) - m^{-1}k_s(a_s) - k_b(a_b) - d(p_s a_s - p_b a_b)$$

The previous version is the special case where $m = 1$. When m increases, this makes subtle antidemocratic action cheaper, i.e., more can be done for the same (direct) cost paid. The FOC for this version are the same other than m^{-1} multiplying the $k_s'(a_s)$ term. Totally differentiating with respect to m gives:

$$v''(a)\frac{\partial a_s}{\partial m} + v''(a)\frac{\partial a_b}{\partial m} - m^{-1}k_s''(a_s)\frac{\partial a_s}{\partial m} + k_s'(a_s)m^{-2} = 0 \tag{7}$$

$$v''(a)\frac{\partial a_b}{\partial m} + v''(a)\frac{\partial a_s}{\partial m} - k_b''(a_b)\frac{\partial a_b}{\partial m} = 0 \tag{8}$$

Solving for $\frac{\partial a_b}{\partial p_s}$ and $\frac{\partial a_b}{\partial p_s}$ gives:

$$\frac{\partial a_s}{\partial m} = \frac{m^{-1}k_s'(a_s)(k_b''(a_b) - v''(a))}{k_s''(a_s)k_b''(a_b) - k_s''(a_s)v''(a) - mk_b''(a_b)v''(a)} > 0$$

$$\frac{\partial a_b}{\partial m} = \frac{m^{-1}k_s'(a_s)v''(a)}{k_s''(a_s)k_b''(a_b) - k_s''(a_s)v''(a) - mk_b''(a_b)v''(a)} < 0$$

Adding these gives $\frac{\partial a_s}{\partial m} > 0$ ■

Combining changes to explain the data Of course, it is possible that more than one of these exogenous parameters can change over time. For example, a combination of changes that could potentially explain some of the observed patterns is (1) an increase in the cost of reported violations, (2) an increase in the effectiveness of subtle violations, and (3) expert coders have the ability to code based on subtle violations which are not reported. Both (1) and (2) can lead to a shift towards subtle undemocratic behavior. However, (1) leads to less aggregate undemocratic behavior while (2) leads to more, which could roughly cancel out, consistent with incumbents and their parties not dominating elections more. Still, it

is not clear why expert coders should rate countries as less democratic given these changes. Perhaps if modifying (3) to assume that expert coders catch *some* violations which are not reported, and (4) an increase in p_s so more reported violations, then expert coders increase in finding subtle violations could capture a combination of real change and more reporting, where the real change is offset by less blatant violations.

All this seems plausible, but this explanation is far less parsimonious than the possibility raised by the coder model that changes in subjective measures simply reflect a change in coding bias.