// NO.21-012 | 02/2021

DISCUSSION PAPER

// KLAUS GRÜNDLER AND TOMMY KRIEGER

# Using Machine Learning for Measuring Democracy: An Update

# Using Machine Learning for Measuring Democracy: An Update.

**Klaus Gründler**[1] **Tommy Krieger**[2]

**February 10, 2021**

**Abstract**

We provide a comprehensive overview of the literature on the measurement of democracy and present an extensive update of the Machine Learning indicator of *Gründler and Krieger* (*2016*, *European Journal of Political Economy*). Four improvements are particularly notable: First, we produce a continuous and a dichotomous version of the Machine Learning democracy indicator. Second, we calculate intervals that reflect the degree of measurement uncertainty. Third, we refine the conceptualization of the Machine Learning Index. Finally, we largely expand the data coverage by providing democracy indicators for 186 countries in the period from 1919 to 2019.

**Keywords:** Data aggregation, Democracy indicators, Machine Learning, Measurement Issues, Regime Classifications, Support Vector Machines

**JEL No.**: C38, C43, C82, E02, P16

[1]LMU Munich, ifo Center for Public Finance and Political Economy, Poschingerstrasse 5, 81679 Munich, Germany. Email: Gruendler@ifo.de

[2]Center for European Economic Research (ZEW), Department of Public Finance, L7 1, 68161 Mannheim, Germany. Email: tommy.krieger@zew.de

# 1 Introduction

Most applied researchers in economics and political science that examine the determinants and consequences of democratic transitions use country-level data, and thus have to choose an indicator that reflects how the level of democracy varies across countries (see e.g. Acemoglu et al., 2008, 2015, 2019, Aidt and Leon, 2016, De Haan and Sturm, 2003, Dorsch and Maarek, 2019, 2020, Gassebner et al., 2013, Harms and Ursprung, 2002, Potrafke, 2012, Zuazu, 2019). Making a well-founded decision on this issue is difficult and requires a great amount of effort because the literature provides various measures of democracy but no clear guidelines that help practitioners to find an index that properly fits to their purposes. The first part of this paper thus gives an extensive overview of the literature on the measurement of democracy. In particular, we describe the methodology of nine well-established democracy indicators and summarize their major strengths and weaknesses in a user-friendly manner. We think that our review fills a gap since existing reviews either require a lot of prior knowledge (see e.g. Knutsen, 2010, Munck and Verkuilen, 2002) or only focus on two or three indices (see e.g. Boese, 2019, Cheibub et al., 2010). Furthermore, in contrast to previous literature surveys, we pay particular attention to the question of how appropriate an index is for applied political economy research.

A recurring problem in existing measures of democracy is that the methods applied for data aggregation are rather simplistic and sometimes even arbitrarily chosen. To address this problem, we developed a Machine Learning approach for measuring democracy in an early paper (see Gründler and Krieger, 2016). The second part of this paper presents an extensive update of our initial index and provides a comprehensive discussion of the underlying method. The large-scale update is necessary for two key reasons: first, our original index has a coverage of only 30 years, and second, its definition of democracy overlaps with other institutional factors. To address these weak spots, we first completely revised the conceptualization of our measure. Instead of assuming a maximalist concept, we now work with a definition that includes only three dimensions: political competition, political participation, and freedom of opinion. Our new concept of democracy thus closely resembles the definition proposed by Dahl (1971). To operationalize our concept of democracy, we compiled ten regime characteristics that are available for 186 countries and the period from 1919 to 2019.

In the course of our update, we did not only improve the conceptualization and the coverage of our democracy indicator but also developed two novel methodological features. The first is that our Machine Learning approach now produces both continuous and dichotomous indices.[1] This novelty is important because the type of scaling that is most appropriate depends on the research question (see e.g. Collier and Adcock, 1999). The second new feature is that we report confidence intervals for our democracy indices. These confidence intervals are valuable since they allow practitioners to take measurement uncertainty into account when conducting empirical analyses.

Compared with other measures of democracy, the particular feature of our index is the Machine Learning approach used for data aggregation. In Gründler and Krieger (2020), we illustrate in great detail the empirical strengths that this

[1] Our Machine Learning indices are publicly available and will be updated on a regular basis. The data can be downloaded from our website: www.ml-democracy-index.net.

approach has compared to the aggregation tools that have often been used for producing continuous democracy indices. In particular, this related paper shows that conventionally used aggregation tools produce indices that are implausibly high (low) for highly autocratic (democratic) regimes. The main reason for this problem is that the applied aggregation tools make relatively strong functional assumptions. When using our Machine Learning approach, we can relax these assumptions and thereby address this issue. In empirical applications, the main consequence of this improvement is that OLS and 2SLS estimates are less likely to be systematically biased.

This paper is structured as follows. Section 2 describes the steps that need to be taken when producing a measure of democracy. Section 3 discusses nine frequently used democracy indicators. Section 4 presents our Machine Learning indices. Section 5 concludes.

# 2 The measurement of democracy

The classical approach for producing a measure of democracy consists of three steps (Munck and Verkuilen, 2002). The first step defines the term "democratic regime" (*conceptualization*). The second step identifies a set of observable regime characteristics, reflecting the key components of the chosen concept of democracy (*operationalization*). The last step establishes a rule that transforms the regime characteristics into a uni-dimensional measure of democracy (*aggregation*). Below, we briefly describe the main challenges that providers of democracy indices face in each of these steps.

## 2.1 Conceptualization

The question of how to define "democracy" is the subject of a longstanding and controversial discussion (for an overview, see Blaug and Schwarzmantel, 2016). Two major challenges exist. The first is to choose the institutional features/dimensions that are associated with "democracy". The other challenge is to specify how these dimensions interact with each other (see Munck and Verkuilen, 2002, Teorell et al., 2019).

Regarding the first issue, the literature broadly distinguishes between three types of concepts: *narrow/minimalist* concepts which mainly care about whether elections for political posts are public and competitive, *realistic* concepts which additionally require universal suffrage and basic political rights, and *broad/maximalist* concepts which also include conditions on other socioeconomic aspects, such as liberty or inequality (O'Donnell, 2001, Munck and Verkuilen, 2002). From a conceptual perspective, all concepts of democracy are equally valid because no objective criteria exist that allow to rank concepts of democracy (Guttman, 1994, Munck and Verkuilen, 2002). From an empirical point of view, however, maximalist concepts might be less appropriate than minimalist or realistic definitions since they conceptually overlap with other economic aspects, such as civil liberties, economic freedom, or the rule of law (Bjørnskov and Rode, 2020, Gutmann and Voigt, 2018).

Two approaches exist on how the dimensions of a concept of democracy can interact with each other (Teorell et al., 2019): the *formative* approach treats each institutional aspect as a necessary condition of democracy, whereas the *reflective* approach assumes that all dimensions are (partial) substitutes and the result of a common factor. We believe that both approaches have their merits and that no clear answer exists on the question of which approach should be used when producing a measure of democracy. We also think that this issue is of great importance for both users and providers of democracy indices. For providers, it matters because several decisions during the aggregation process crucially depend on whether the chosen concept of democracy is formative or reflective (for details, see Section 2.3). For users, this conceptual decision plays a role because interpretations of empirical results must be consistent with the underlying conceptual assumptions.

## 2.2 Operationalization

Most of the institutional features that are typically part of standard concepts of democracy are not directly measurable. A common way to address this issue is to create one (or more) expert-based sub-indices for each of the chosen dimensions of democracy. The merit of this approach is that subjective scores can (at least in principle) be built for any regime and any period. However, the flip side is that coders might be biased due to personal experiences or interests and that subjective assessments might considerably differ between coders. Munck and Verkuilen (2002) thus recommend to complement expert-based sub-indicators with objective regime characteristics. The main problem of using objective information is that this data does usually not exist for all regimes and periods.

## 2.3 Aggregation

In mathematical terms, data aggregation requires to specify a function ($f$) that maps a set of observable regime characteristics ($\mathbf{z}$) onto a level of democracy ($d$):

$$d_{it} \;=\; f(\mathbf{z}_{it}) \qquad \forall\, \mathbf{z}_{it} \;=\; \left(z_{it}^{1},\; \ldots\;,\; z_{it}^{r}\right) \tag{1}$$

where $i$ denotes a country and $t$ a particular point in time (Gründler and Krieger, 2016).

Munck and Verkuilen (2002) suggest that providers of democracy indices must meet two main challenges during the aggregation process. The first is that they must decide on the scale of the index. Four types of scaling exist: *dichotomous* scales, *ordinal* scales, *graded/quasi-continuous* scales, and *continuous* scales. All of these scales have their strengths and weaknesses. While indices with dichotomous or ordinal scales often perform better from a conceptual perspective, continuous and graded indicators have some advantages from an empirical perspective because of their greater discriminating power (for further details, see Sections 3 and 4).

The other major challenge is to specify the aggregation function. This task is difficult for two reasons: first, the form of the aggregation function needs to be consistent with the conceptual assumptions, and second, neither the true level of democracy nor the shape of the aggregation function are directly observable. The conventional way to deal with the latter issue is to make specific assumptions about the functional relationship between the observable regime characteristics and the level of democracy. This approach is, however, often subject to criticism, in particular when the democracy index has a continuous or graded scale (see e.g. Cheibub et al., 2010, Munck and Verkuilen, 2002). A key reason for this criticism is that the chosen functional forms are usually quite simplistic and not grounded in theory.

A common belief among users of democracy indices is that the methodological details of the aggregation process are of little practical importance. Gründler and Krieger (2020) challenge this view by showing theoretically and empirically that the choice of the scale as well as the choice of the aggregation technique significantly affects the size of regression coefficients when using a democracy indicator as a dependent variable in an OLS or 2SLS regressions. In particular, scaling decision have an effect on the amount of measurement uncertainty and thus on the size of the attenuation bias. Furthermore, changes in the aggregation function of a continuous index cause systematic changes in the index values of regimes at the lower and/or upper end of the democracy spectrum, which in turn affects the magnitude of both OLS and 2SLS estimates.

# 3 Popular democracy indicators

This section describes the methodologies, strengths, and weaknesses of nine widely used measures of democracy (for a summary, see Table B.9). Our motivation is twofold: First, we think that the literature lacks a paper that summaries the main aspects of standard indices in an user-friendly manner.[2] Second, we can discuss and motivate our Machine Learning indicator (see Section 4) in a clearer way when having previously established the major strengths and weaknesses of existing democracy indices.[3]

## 3.1 Boix-Miller-Rosato Indicator

### 3.1.1 Methodology

The *Boix-Miller-Rosato* (*BMR*) indicator is a dichotomous democracy index that labels a regime as democratic (1) if it met the following conditions (Boix et al., 2013):

- The majority of male citizens was eligible to vote.[4]

[2]The question that implicitly underlies our literature review is how useful a particular measure of democracy is for empirical analyses. A legitimate question in this regard is whether the existing measures should be evaluated from this perspective. We think that this is indeed the case since democracy indices have been used in numerous empirical studies as a dependent or independent variable. However, we do not believe that the empirical applicability is the only dimension based on which democracy indices can be evaluated. The reason for why we pay attention to this aspect is that an extensive discussion of the relevant theoretical questions regarding the conceptualization/operationalization of democracy was already provided by Munck and Verkuilen (2002). Our review will therefore complement rather than substitute Munck and Verkuilen's seminal work.

[3]As outlined in greater detail in Section 4, the basic reason for why we started to develop a Machine Learning approach for measuring democracy was that we aimed to find an approach that addresses the empirical problems that have been identified for existing democracy indices. To understand the strengths of our method, it is thus of great importance to illustrate the methodological advantages and disadvantages of existing indicators.

[4]In their latest update, the providers of the BMR index present a second measure of democracy which additionally requires that the majority of female citizens had the right to vote in national elections (Boix et al., 2018).

- The executive power was directly or indirectly elected in popular elections and is responsible to the voters or to a legislature.
- The legislature (and when directly elected the executive) was chosen in free and fair elections.

### 3.1.2 Discussion

The most recent version of the BMR database includes democracy indices for 222 countries in the period from 1800 to 2015 (Boix et al., 2018). In terms of data coverage, the BMR indicator is thus among the three best performing measures of democracy. A second major strength of the BMR indicator is its methodological consistency: Boix et al. (2013) first argue that the three institutional aspects that they take into account when classifying regimes constitute necessary conditions for democracy, and then apply operationalization and aggregation approaches that fit together with their conceptual assumptions. Finally, the definition of democracy underlying the BMR indicator only includes requirements on political competition and political participation. This rather narrow conceptualization is notable since conceptual overlaps with other institutional features (such as the rule of law or economic freedom) are thus unlikely when applying the BMR index in a regression analysis on the causes and consequence of democracy.

An objection against all dichotomous measures of democracy (and thus also against the BMR index) is that these indicators lack discriminating power (Bollen and Jackman, 1989). From an empirical perspective, a low degree of differentiation creates problems, especially when the relationship between democracy (or one of its dimensions) and an economic outcome is non-linear. For example, the BMR index does not differentiate between regimes in which all male citizens have the right to vote and regimes in which male suffrage is restricted. This implies that the BMR index can hardly be used for examining how government expenditures respond to institutional changes because some theories suggest that a U-shaped relationship might exist between suffrage and public spending on human capital (see e.g. Aidt et al., 2010). Another general objection against dichotomous democracy indices is that transition years can often not precisely be identified because many political systems change gradually.

Boix et al. (2013) define that regimes require "free" and "fair" elections to be democratic. Operationalizing this condition is difficult since objective criteria for "free" and "fair" do not exist. Consequently, all coding decisions for this crucial condition are based on subjective evaluations. For many cases, such a subjective assessment is uncontroversial. For the other cases, uncertainty measures should be provided, according to Lührmann et al. (2018). The simple way of indicating the existence of measurement uncertainty is to create a dummy variable that shows whether coders are unsure about their assessment. The more sophisticated way is to report a confidence interval that reflects the level of measurement uncertainty. Boix et al. (2013) applies neither of these two approaches. Thus, testing whether the findings of an empirical analysis depend on the handling of borderline countries is impossible when using the BMR index.

Munck and Verkuilen (2002) prompt providers of democracy indicators to publish both their final index and the underlying raw data. Boix et al. (2013) only partly comply with this request since their published database does not include their raw data. Besides transparency issues, this incomplete provision is problematic since users of the BMR measure can neither check whether their results depend on the choice of the aggregation method nor whether they are caused by a specific aspect of democracy.

## 3.2 Democracy-Dictatorship indicator

### 3.2.1 Methodology

The *Democracy-Dictatorship* (*DD*) indicator is a binary measure of democracy that was initially proposed by Alvarez et al. (1996) and later updated by Cheibub et al. (2010) and Bjørnskov and Rode (2020). The DD indicator classifies a regime as democratic (1) only if it satisfies the following four conditions:

- The head of government was elected by a popular vote or by a popularly elected body.
- The members of the legislature were elected through a popular election.
- More than one party attended the parliamentary elections.
- An alternation in power took place under the same election rules that brought the incumbent into power.

### 3.2.2 Discussion

The latest version of the DD index covers the period from 1950 to 2020 and is available for 192 sovereign countries, 16 self-governing territories, and 96 colonies (Bjørnskov and Rode, 2020). The total coverage of the DD index is thus smaller than the coverage of the BMR index. For the Post-World-War-II period, however, the DD index has two major advantages over the BMR index: first, it distinguish between different types of colonies, and second, publicly available information exists that indicates which condition(s) a country violates that is not classified as a "democracy".[5] The latter feature is important not only for transparency reasons, but also because it allows users to investigate whether their results are triggered by a particular aspect of democracy. A merit that the DD index shares with the BMR index is the consistency between the definition of democracy and the aggregation procedure.

Besides the general objections against dichotomous measures regarding their low discriminating power, the DD index is subject to criticism mainly for conceptual reasons. Munck and Verkuilen (2002) suggest that the concept underlying the DD indicator is too minimalistic since it does not include any requirement on political participation. We agree with the view of Munck and Verkuilen (2002), but it is important to mention that Bjørnskov and Rode (2020) add information on suffrage rules to the DD database. Users of the DD indicator can thus simply extend the definition of democracy to examine whether their regression results are robust to conceptual changes. A second major conceptual problem of the DD index is the condition that requires an alternation in power. Knutsen and Wig (2015) illustrate that the DD indicator systematically misclassifies "young" democracies due to this

[5] Most providers of democracy indicators classify colonies as non-democratic regimes because their governments could not freely decide on laws and policy measures.

condition.[6] A frequent consequence of this misclassification is that point estimates are biased when using the DD index as the explanatory variable in a regression analysis on the effects of institutional transitions (for details, see Knutsen and Wig, 2015).[7]

## 3.3 Polity IV indicator

### 3.3.1 Methodology

The *Polity IV* index is a quasi-continuous/graded democracy indicator that ranges from $-10$ (most autocratic) to $+10$ (most democratic). To create this indicator, Marshall et al. (2019) proceed in three steps: First, they assume a definition of democracy that consists of five institutional aspects:

- Constraints on chief executive.
- Competitiveness of the chief executive recruitment.
- Openness of the chief executive recruitment.
- Regulation of participation.
- Competitiveness of participation.

Afterwards, they evaluate each country with more than 500,000 inhabitants with regard to each of the five institutional aspects. Finally, they add together the five sub-indicators developed in the previous step.

### 3.3.2 Discussion

The Polity IV indicator is available for the period from 1800 to 2018 and is annually updated. The major strength of this measure of democracy is that it has both a great coverage and substantial discriminating power. Furthermore, the five sub-indicators are publicly available. Users can thus check whether their estimation results are driven by particular aspects of democracy or depend on the aggregation method.

Many conceptual and methodological concerns exist against the Polity index (for extensive discussions, see Boese, 2019, Cheibub et al., 2010). From a conceptual perspective, two of them are particularly relevant for practitioners. First, the concept of democracy underlying the Polity index overlaps with other institutional factors. For example, the dimension "constraints on chief executive" takes into account the independence of the courts (Marshall et al., 2019), which in turn is a component of most indices measuring the rule of law (Gutmann and Voigt, 2018). Second, the providers of the Polity indicator suspend their coding procedure if a regime has no or a transitional government or if it is occupied by another country. To fill the data gaps that originate from these suspensions, Marshall

[6] For example, post-Apartheid South Africa is classified as an autocracy even though free and fair elections have been regularly held since 1994. The reason for this classification is that the African National Congress (ANC) won all national elections since the enfranchisement of black citizens in 1994.

[7] Alvarez et al. (1996) admit that the alternation rule is likely to create a systematic error in their classification.

**Table 1** Freedom House Indicator: Creation of PR and CL index.

| | +1 | +2 | +3 | +4 | +5 | + 6 | +7 |
|---|---|---|---|---|---|---|---|
| PR index | 36 – 40 | 30 – 35 | 24 – 29 | 18 – 23 | 12 – 17 | 6 – 11 | 0 – 5 |
| CL index | 53 – 60 | 52 – 44 | 35 – 43 | 26 – 35 | 17 – 25 | 8 – 16 | 0 – 7 |

**Notes**: This table shows how Freedom House transforms the sum of the expert-based regime characteristics on political rights (PR) and civil liberties (CL) into graded indicators with seven-tiered scales.

et al. (2019) propose to assign the value of ±0 to such a regime. We argue that this simple interpolation procedure is problematic since it often creates spurious changes in the level of democracy.[8]

Several discussants of the Polity index criticize that the 21 categories overlap, especially at the center of the spectrum (see e.g. Gleditsch and Ward, 1997). We share this view, but also think that classifying hybrid regimes without having any measurement uncertainty is impossible. However, in contrast to other indices, the Polity database does not provide confidence intervals that indicate the extent of measurement uncertainty. A second major methodological problem of the Polity index is the aggregation rule. Munck and Verkuilen (2002) argue that weighting all regime characteristics equally is inappropriate because some of them are redundant. Gründler and Krieger (2020) and Teorell et al. (2019) show that the application of an additive rule creates doubtful classifications at the lower end of the spectrum. Cheibub et al. (2010) and Treier and Jackman (2008) conclude that the aggregation procedure used to compute the Polity index lacks any theoretical foundation.

## 3.4 Freedom House Indicator

### 3.4.1 Methodology

*Freedom House* (*FH*) publishes two democracy indices. The first index is graded/ quasi-continuous and ranges from +2 (most democratic) to +14 (most autocratic), while the second is ordinal and consists of three categories (*free*, *partly free*, *not free*). Below, we will mainly focus on the first index since the ordinal measure is just a simple extension of the graded index.[9]

FH distinguishes between two aspects of democracy: political rights (PR) and civil liberties (CV). For each of the two, FH creates a separate index, ranging from +1 to +7. The final FH indicator is the sum of the PR and the CL indicator. To produce the PR/CL index, FH takes 10/15 regime characteristics into account. All of them are expert-based and vary from 0 to +4. The PR/CL index of a regime depends on the sum of the 10/15 regime characteristics. In Table 1, we show the allocation scheme (for further details, see Freedom House, 2019).

[8] For example, when applying this procedure, the Polity index of Poland increases from −6 to 0 in 1939 when Poland was occupied by Nazi Germany.

[9] A country is classified as "free" when the graded FH index does not exceed +5, as "partly free" when the graded FH index is larger than +5 and at most +10, and "not free" when the graded FH index exceeds +10.

### 3.4.2 Discussion

The FH indices exist for almost all independent countries of the world and most self-governing territories. However, the time spans of the FH measures are relatively short since they are only available from 1972 onward. Compared to the Polity IV indicator, a clear advantage of the FH indicators is that spurious changes due to anarchy or foreign occupation do not exist. Furthermore, FH provides much more information about the institutional aspects that the coders take into account when evaluating countries.

From a practical perspective, a key problem of the FH indicators is that their underlying concept of democracy is maximalist and thus overlaps with many other concepts (Munck and Verkuilen, 2002). Some scholars try to address this issue by using only the PR index. We agree that this approach alleviates the problem to some extent since the CL index includes most of the institutional aspects that are not part of a narrow or realistic definition of democracy. However, we also believe that using only the PR index does not fully solve the problem since this measure takes political corruption into account.

The FH indicators share the basic methodological weaknesses of the Polity IV index (i.e. no confidence intervals, overlapping categories, non-theorized aggregation rule). In addition to them, the FH indices have two further weak spots: first, all threshold values are completely arbitrary, and second, users can hardly investigate whether their results depend on the aggregation rule because the 25 expert-based regime characteristics are publicly available only from 2006 onward (Boese, 2019, Cheibub et al., 2010).

## 3.5 Acemoglu-Naidu-Restrepo-Robinson Indicator

### 3.5.1 Methodology

Acemoglu et al. (2019) start their work with the claim that all democracy indices indicate spurious regime transitions and argue that this issue can be considerably alleviated when creating a dichotomous measure that combines the ratings of the Polity, FH, DD, and BMR indicator according to the following rules:

- A regime is "democratic" when the Polity index is greater than 0 and the FH index does not exceed 10.
- A regime is "democratic" when the Polity index does not exist, the FH index does not exceed 10, and either the DD or the BMR index indicates "democratic".
- A regime is "democratic" when the FH index does not exist, the Polity index is greater than 0, and either the DD or the BMR index indicates "democratic".
- A regime is "democratic" when neither the Polity index nor the FH index exist, and the other two indices indicate "democratic".
- A regime is "non-democratic" when none of the former four cases applies.

### 3.5.2 Discussion

The *Acemoglu-Naidu-Restrepo-Robinson* (*ANRR*) database includes indices for 184 countries and the period form 1960 to 2010. Compared to most other popular democracy indicators, a key advantage of the ANRR index is that all underlying measures are publicly available and are updated on a regular basis. Users can thus easily extend the time span of the ANRR index if this is needed for their research project.

We argue that the ANRR indicator has severe shortcomings. First of all, the underlying concept of democracy is not the same for all country-year observations. A reason for this inconsistency is that FH defines "democracy" in a much broader way than the providers of the other indices (for details, see Sections 3.1 – 3.4) has only been publishing indices since 1972. Second, since the Polity and the FH index conceptually overlap with other institutional factors (see Sections 3.3 and 3.4), the ANRR index does as well. Third, Acemoglu et al. (2019) themselves indicate that their classification method creates spurious regime changes. To address this issue, Acemoglu et al. (2019) manually recode all transitions that are not plausible from their perspective.[10] We believe that these reclassifications are problematic since no selection criteria exist that clearly suggest when a recoding is necessary. Fourth, uncertainty measures do not exist. Finally, the thresholds defined for the FH and the Polity index lack a solid theoretical foundation and are thus arbitrary (see Bogaards, 2012).

## 3.6 Vanhanen index

### 3.6.1 Methodology

The measure of democracy by Vanhanen (2000) has a continuous scale and builds upon a definition that considers political participation and political competition as necessary and sufficient conditions for democracy. Vanhanen (2000) operationalizes these two institutional features with two objective variables: the voter-population-ratio and the share of votes that was not won by the strongest party/candidate.[11] The *Vanhanen index* is the product of these two regime characteristics.

### 3.6.2 Discussion

Vanhanen's indicator is available for more than 190 countries and the period from 1810 to 2018. Updates of the index are not provided annually, but on a regular basis. The regime characteristics are freely accessible. Users can thus easily test whether their findings depend on the aggregation technique.[12] Another remarkable feature is the comprehensive documentation file which lists (by country and year) the used data sources.

[10] For example, when applying the basic procedure of Acemoglu et al. (2019) without manual adjustment, the ANRR index suggests that South Africa moved from autocracy to democracy in the mid 1980s. Acemoglu et al. (2019) argue that this transition is spurious since the Apartheid regime was in place until 1994.

[11] To account for the constitutional differences across democratic regimes, Vanhanen (2000) weights presidential and legislative elections according to their relevance for the political decision making process.

[12] Vanhanen (2000) admits that his aggregation method cannot fully be grounded in theory and thus recommends this kind of robustness check.

Munck and Verkuilen (2002) criticize the Vanhanen index especially due to the two regime characteristics used to operationalize political competition and political participation and the way by which Vanhanen (2000) aggregates them. The main problem is that regimes are only classified as fully democratic by a multiplicative indicator when all underlying regime characteristics reach their highest level. With regard to the Vanhanen indicator, this implies that parliaments need to be highly fractionalized and that turnout rates have to be 100 percent in fully democratic countries. We argue that these requirements are excessive. For example, if voters voluntarily abstain from voting, indicating a lack of democracy is not appropriate from our perspective.

A second major problem of the Vanhanen indicator is that precise information on the two objective regime characteristics are not available for all country-year observations. To address this issue, Vanhanen (2000) applies various interpolation procedures. We think that this is an adequate way for dealing with missing raw data. However, we also believe that providers of objective indices should publish uncertainty measures if their data is incomplete. For the Vanhanen index, such measures do not exist.

## 3.7 Unified Democracy Score

### 3.7.1 Methodology

The *Unified Democracy Score* (*UDS*) is a continuous indicator that was initially developed by Pemstein et al. (2010). The basic idea behind the UDS is to create a new measure of democracy out of ten existing indices, using a Bayesian latent variable approach. The list of underlying indices includes: the DD, the FH, the Polity, and the Vanhanen index, as well as the measures proposed by Arat (1991), Bollen (1999), Bowman et al. (2005), Coppedge and Reinicke (1990), Hadenius (1992), and Gasiorowski (1996).[13]

### 3.7.2 Discussion

The first version of the UDS covers the 1946-2000 period and exists for almost all countries of the world (Pemstein et al., 2010). A recent update by Márquez (2018) improves the coverage such that the UDS is now available from 1815 onward. A major strength of the approach used by Pemstein et al. (2010) is that it does not only produce an index for each country-year observation, but also a measure that indicates the extent of measurement uncertainty.[14]

According to Pemstein et al. (2010), a second strength of their Bayesian latent variable approach is that it can combine various measures of democracy although their coverage rates notably differ. However, Gründler and Krieger (2016) show that using this methodological feature comes at high costs since it creates conceptual inconsistencies and spurious changes in the estimated level of democracy when the

[13] In our review, we do not discuss the indices by Arat (1991), Bollen (1999), Bowman et al. (2005), Coppedge and Reinicke (1990), Hadenius (1992), and Gasiorowski (1996) because they are hardly used anymore.

[14] The confidence intervals of the UDS are largest for highly autocratic/democratic regimes and shortest for hybrid regimes. Teorell et al. (2019) argue that this pattern is counterintuitive. We share this view and thus believe that the confidence intervals of the UDS should be interpreted with caution.

number of underlying indicators changes. Furthermore, the Bayesian latent variable approach often creates implausibly high/low scores for highly autocratic/democratic regimes. An empirical consequence of this problem is that regression results are systematically biased (Gründler and Krieger, 2020).

## 3.8 Lexical Index of Electoral Democracy

### 3.8.1 Methodology

The *Lexical Index of Electoral Democracy* (*LIED*) by Skaaning et al. (2015) is an ordinal and expert-based measure that distinguishes between the following types of political regimes.

- Regimes without elections.
- Regimes with no-party or one-party elections.
- Regimes with multiparty parliamentary election.
- Regimes with multiparty parliamentary and presidential elections.
- Regimes with minimally competitive multiparty election for legislature and executive.
- Regimes with minimally competitive multiparty election and full male or female suffrage for legislature and executive.
- Regimes with minimally competitive multiparty election and full male and female suffrage for legislature and executive.

### 3.8.2 Discussion

The index proposed and regularly updated by Skaaning et al. (2015) starts in 1789 and provides information on all independent countries of the world. LIED is thus among the democracy indicators with the best coverage. Furthermore, due to its clear coding rules, its narrow definition of democracy, and its theoretically well grounded aggregation procedure, we believe that LIED can be hardly criticized from a conceptual or methodological perspective.

Despite its great merits, we have some doubts whether LIED can be effectively used in applied studies. The key problem is its ordinal scale. The correct way of using LIED in a regression analysis in which the level of democracy serves as an explanatory variable is to create a dummy variable for each of the seven types of regimes. Of course, this is unproblematic in an OLS analysis. However, an OLS estimate can usually not be causally interpreted. Many scholars thus also present results from instrumental variable regressions. Running this type of regression with the measure by Skaaning et al. (2015) is virtually impossible since one instrumental variable would be necessary for each regime type.[15]

[15] An inappropriate approach to address this problem is to treat LIED as a graded democracy indicator and to add it as a single variable to the regression model. The reason for why this approach does not work is that the degree of democratic improvement is not always the same when switching from one regime type to the next. For example, moving from a regime without elections to a regime with single-party elections hardly increases the degree of democratization

**Table 2** V-Dem's Polyarchy index: Regime Characteristics.

| Description | Type | Sub-indicators | Concept | Aggregation | Weight |
|---|---|---|---|---|---|
| Extent to which chief executive and legislature are elected | Expert-Based | 16 | Formative | Manual | 0.25 |
| Associational Autonomy | Expert-based | 6 | Reflective | Bayesian item responds model | 0.25 |
| Free and fair elections | Expert-based | 8 | Reflective | Bayesian item responds model | 0.125 |
| Freedom of expression and alternative sources of information | Expert-based | 9 | Reflective | Bayesian item responds model | 0.25 |
| Suffrage | Objective | 1 | - | - | 0.125 |

**Notes**: This table lists the five regime characteristics that constitute the Polyarchy index by Teorell et al. (2019).

## 3.9 V-Dem's Polyarchy Index

### 3.9.1 Methodology

Varieties of Democracy (V-Dem) is an international research project that aims to provide institutional information on all countries of the world. In total, the V-Dem database includes more than 450 regime characteristics and five democracy indices (Coppedge et al., 2019). Since these indicators differ mainly with regard to their underlying concept of democracy, we do not discuss all of them, but focus on the *Polyarchy index* by Teorell et al. (2019). We choose this index because the other democracy indicators in the V-Dem database are only extensions of the Polyarchy index.

V-Dem's Polyarchy index is a continuous measure of democracy that takes five regime characteristics into account (see Table 2). Four of them consist of various expert-based sub-indicators. The procedure used to aggregate these sub-indicators depends on whether a regime characteristic is reflective or formative (for further details, see Teorell et al., 2019).

Teorell et al. (2019) apply a two-step procedure to transform their five regime characteristics into the Polyarchy index. In the first step, they create an additive and a multiplicative index (for the weighting scheme, see Table 2). In the second step, Teorell et al. (2019) compute the Polyarchy index by taking the mean of the additive and the multiplicative indicator.

### 3.9.2 Discussion

V-Dem's Polyarchy index is available for 202 countries in the period from 1789 to 2019. The underlying definition of democracy is realistic and thus hardly overlaps with other institutional aspects. In addition, the selection and construction of the regime characteristics is well grounded in theory. Another notable feature is that Teorell et al. (2019) compute a confidence interval for each rating. All raw data is

because political competition still does not exist. When instead switching from single-party to multiparty elections, the increase in the degree of democratization is remarkable because a key condition of democracy is then at least partly satisfied.

publicly available and the code book of the V-Dem database clearly describes their coding rules and aggregation methods.

We believe that the aggregation procedure is the only weak spot of V-Dem's Polyarchy indicator. Two problems are notable. First, Teorell et al. (2019) do not provide a clear reasoning for why their additive and their multiplicative indicator should be equally weighed and additively combined. Second, Gründler and Krieger (2020) illustrate that combining an additive and a multiplicative index does not suffice to eradicate the methodological weaknesses of additive and multiplicative indicators.

### 3.10 Recommendations

From a practitioner's point of view, Sections 3.1 – 3.9 might raise the question which of the nine aforementioned democracy indicators should be applied in an empirical investigation. Responding to this question in a general and profound manner is difficult since each index has its advantages and disadvantages and it often crucially depends on the research design and the research question how problematic a particular weak point actually is. For instance, the ordinal scale of the Lexical Index of Electoral Democracy (see Section 3.8) is unproblematic when running OLS regressions but cause empirical problems if a 2SLS approach is planned as the identification strategy.

To find a measure of democracy that fits well to their research project, we recommend researchers to ask themselves three basic questions and to use the principle of exclusion. The first basic question is which of the available indices creates a conceptual overlap with the factor whose relationship to democracy is analyzed in the research project. For example, if a scholar aims to show how political institutions affect economic institutions, he/she cannot use a democracy indicator that takes into account aspects such as judiciary independence or the rule of law. From our perspective, such conceptual overlaps are especially likely when using measures with broad definitions such as the Freedom House indices (see Section 3.4), the Unified Democracy Score (see Section 3.7), and the ANRR index (see Section 3.5). Second, practitioners can restrict their list of potential indices when deciding whether they assume a formative or a reflective concept (for details, see Section 2.1). Finally, we think that researchers should become clear about whether a continuous or dichotomous democracy indicator is more appropriate for their analysis (for details, see Section 2.3 and Collier and Adcock, 1999).

Of course, even if scholars answer the three above-mentioned questions, it is possible that they are left with more than one democracy indicator. In such a case, we recommend studying the properties of the eligible indicators in some detail. In particular, when a decision needs to be made between different binary indicators, we argue that those measures that are based on arbitrary thresholds (see e.g. the ANRR index) should not be chosen because the selection of the threshold shapes the regression results (see Bogaards, 2012, Gründler and Krieger, 2020). If the shortlist includes several continuous indices, we believe that it is important to check the aggregation procedures since the aggregation method has a significant effect on the magnitude of the regression coefficients (Gründler and Krieger, 2020). In any case, we think that researchers should present robustness checks in which they illustrate how their results change if they replace their preferred democracy indicator with another index that is appropriate for their research question.

# 4 Machine Learning Indicator

## 4.1 Motivation

We started our project from the observation that many existing measures of democracy are subject to severe criticism due to their aggregation methods. A frequently lamented weak point is that the assumptions about the functional relationship between the regime characteristics and the level of democracy are arbitrary and too simplistic. Our main objective is to address these issues. To this end, we will present and discuss a Machine Learning approach that solves highly non-linear optimization problems to identify an appropriate aggregation function.

## 4.2 Support Vector Machines

Support Vector Machines (SVM) is a supervised Machine Learning technique for pattern recognition that aims to reveal the unknown relationship between a set of input characteristics and an outcome variable (Steinwart and Christmann, 2008). In comparison to conventionally used statistical methods (e.g. OLS or GMM), a key feature of SVM is that one does not have to explicitly specify the function that maps the inputs onto the output. SVM is thus a particularly useful tool if the unknown functional relationship is complex (Bennett and Campbell, 2000, Guenther and Schonlau, 2016).[16]

To compute democracy indices, we will use a SVM classification and a SVM regression tool.[17] The key requirement for using these tools is to have a set of observations for which we observe both the inputs and the output. In our case, meeting this prerequisite is not trivial because democracy is often considered as a latent variable. In Section 4.3.1, we will argue that the level of democracy is observable for some regimes and that we can exploit these regimes to reveal the functional relationship between a set of regime characteristics and the degree of democratization.

## 4.3 Methodology

We now provide a general description of how we use SVM methods to produce measures of democracy. To this end, we assume that we have a sample $\mathcal{Z}$ that includes $n \gg 0$ country-year observations $(i, t)$ for which we observe the regime characteristics $z^1, \ldots, z^r$. Below, we refer to this sample as *priming data*.

[16]Because of its high prediction accuracy in complex situations, SVM tools have been used in a variety of research fields. For example, medical scientists use SVM tools to classify cancer cells (see e.g. Guyon et al., 2002).

[17]Appendix A provides a short description of the mathematical foundations of Support Vector Regressions and Support Vector Classifications. For details, see Abe (2005) and Steinwart and Christmann (2008).

### 4.3.1 Priming data

Applying SVM tools for the measurement of democracy requires a set of regimes (*priming data*) based on which the unknown functional relationship between the regime characteristics and the level of democracy can be learned. Satisfying this prerequisite is difficult because regimes are suited only when we can directly (i.e. without knowing the true aggregation function) observe their level of democracy,

In the political science literature, it is well-established that the classification of regimes whose levels of democracy lie at the ends of the dictatorship-democracy-spectrum is uncontroversial (see Cheibub et al., 2010, Lindberg et al., 2014).[18] For example, the vast majority of social scientists agrees that Nazi Germany was an autocracy, whereas today's Germany is widely considered a democratic regime. We argue that a negligibly low degree of disagreement about how to classify a regime implies that its level of democracy is directly observable. We are therefore also convinced that uncontroversial regimes are an appropriate choice for building the priming data.

### 4.3.2 Aggregation procedure

For a given set of priming data $\mathcal{P}$, we proceed in five steps to transform the regime characteristics into a measure of democracy. In the first step, we decide on the scale of the democracy index. The existing options are: i) a dichotomous scale, or ii) a continuous scale that ranges from 0 (fully autocratic) to 1 (fully democratic). In the second step, we randomly select the training set $\mathcal{T}_s$ out of the priming data $\mathcal{P}$.[19] In the third step, we use the training set $\mathcal{T}_s$ and a SVM method to estimate the aggregation function:

$$f_s\colon \mathbb{R}^r \to \mathcal{D} \qquad \text{where} \qquad \mathcal{D} = \{0, 1\} \qquad \text{or} \qquad \mathcal{D} = [0, 1].$$

The choice of the SVM technique thereby depends on the scaling decision taken in the first step. When producing a dichotomous measure, we apply the SVM classification tool (see Appendix A.1). Otherwise, we use the SVM regression tool (see Appendix A.2).[20] In the fourth step, we use the aggregation function $f_s$

[18] Lindberg et al. (2014) literally write that "*almost everyone agrees that Switzerland is democratic and North Korea is not.*" In a similar vein, Cheibub et al. (2010) point out that "*no measure will produce very different ratings of political regimes in, say, the United Kingdom, Sweden, North Korea or Sudan.*"

[19] We use a uniformly distributed random number generator to determine the total number of observations in the training set $\mathcal{T}_s$. The size of the training set is between 20 and 50 percent of the size of the priming data set:

$$|\,\mathcal{T}_s\,| \sim \mathfrak{U}(\,0.2 \cdot |\,\mathcal{P}\,|,\, 0.5 \cdot |\,\mathcal{P}\,|)$$

Note that changing the possible size of the training set has no notable effect on the final democracy indicators (results available upon request).

[20] A few methodological choices are necessary when using these SVM tools. Both methods require the choice of a penalization parameter $C$ and a kernel function (see Appendix A). We set the penalization parameter $C$ equal to 1 as proposed by Mattera and Haykin (1999) and apply the conventional Gaussian RBF kernel as recommended by Guenther and Schonlau (2016). Using the SVM regression tool additionally requires to choose a margin parameter $\varepsilon$ and labels for the observations in the priming data. We set the margin parameter to 0.025 and set the labels for autocratic/democratic regimes in the priming data to 0.025/0.975. Our parameter choice implies that the degree of democratization of autocratically (democratically) labeled regimes can vary between 0 (0.95) and 0.05 (1) without penalization.

to compute the democracy index

$$d_{it_s} = f_s(\mathbf{z_{it}}) = f_s(z_{it}^1, \dots, z_{it}^r)$$

for each country-year observation $(i, t)$ in sample $\mathcal{Z}$. In the last step, we repeat the process from step 2 to step 4 for all iterations $s \in \{0, \dots, s_{max}\}$.[21]

### 4.3.3 Output

The aggregation tool described in the previous section produces three outputs. The first is a continuous or dichotomous measure that reflects the level of democracy of any country-year observation in the sample. We create this index by taking the median of the $s_{max}$ indices estimated during the aggregation procedure:

$$d_{it} = \text{med}\left[d_{it_s}\right]_{s \in \{0, \dots s_{max}\}} = \text{med}\left[f_s(\mathbf{z_{it}})\right]_{s \in \{0, \dots s_{max}\}}.$$

The second output is the standard deviation of the $s_{max}$ indices:

$$\sigma_{it} = \text{sd}\left[d_{it_s}\right]_{s \in \{0, \dots s_{max}\}} = \text{sd}\left[f_s(\mathbf{z_{it}})\right]_{s \in \{0, \dots s_{max}\}}.$$

The third output is a distribution of indices that consists of the percentiles of the $s_{max}$ indices:

$$\Delta_{it} = \{d_{it}^0, d_{it}^1, \dots, d_{it}^{99}, d_{it}^{100}\}$$

where $d_{it}^j$ denotes the $j$-th percentile.[22]

## 4.4 Implementation

To produce an SVM indicator in practice, we need to compile a set of regime characteristics and have to choose the priming data. In this section, we briefly describe how we addressed these issues in our initial work and describe how we improved the democracy index that we published a few years ago. We also give a description of how we revised our index in order to eradicate the drawbacks of the initial Machine Learning Democracy indicator.

### 4.4.1 Initial setup

In Gründler and Krieger (2016), we used our SVM approach to create continuous democracy indicators for 185 countries in the period from 1981 to 2011. The underlying regime characteristics came from different sources, including Cingranelli et al. (2014), Freedom House (2019), Gibney et al. (2013), and Vanhanen (2000), and captured a variety of institutional factors. To select country-year observations that can serve as priming data, we used the Polity IV index and labeled a regime as democratic (autocratic) if its Polity score was equal to +10 (smaller than −7).

While solving the problem of arbitrary aggregation, our initial Machine Learning democracy indicator has three weak points. The first is its relatively low temporal

[21] In our baseline specification, we set $s_{max} = 2000$. This choice ensures that our aggregation procedure converges. The numerical evidence is available upon request.

[22] In Section 4.5 and 4.6, we will show that the second and third output of our aggregation procedure are useful to reflect the extent of measurement uncertainty.

**Table 3** Machine Learning democracy indices: Regime Characteristics.

| Dimension | Regime Characteristics | Type | Description |
|---|---|---|---|
| Political Participation | Part I | Objective | Share of adult citizens with legally granted suffrage. |
| | Part II | Objective | Ratio between number of voters and number of eligible voters. |
| | Part III | Objective | Ratio between number of voters and number of inhabitants. |
| Political Competition | Comp I | Subjective | Measure of party pluralism. |
| | Comp II | Objective | Share of votes not won by the strongest party/candidate. |
| | Comp III | Objective | Share of parliamentary seats not won by the strongest party. |
| | Comp IV | Objective | Ratio between share of votes won by runner-up party/candidate and share of votes won by strongest party/candidate. |
| | Comp V | Objective | Ratio between share of parliamentary seats won by runner-up party and share of parliamentary seats won by strongest party. |
| Freedom of Opinion | FreeOp I | Subjective | Measures freedom of discussion of male citizens. |
| | FreeOp II | Subjective | Measures freedom of discussion of female citizens. |

**Notes**: This table lists the ten regime characteristics that constitute our Machine Learning democracy indicators.

coverage. The second key weakness is its concept of democracy, which is relative broad and thus overlaps with some other factors such as economic freedom or the rule of law. Finally, some of the regime characteristics are expert-based ratings that have often been criticized because of their aggregation procedures. When using them, we thus face the risk that their methodological problems cause biases in our Machine Learning democracy index.

### 4.4.2 Revision: conceptualization and operationalization

To address the weak spots of the initial Machine Learning democracy index, we completely revised its conceptualization and operationalization. Our concept of democracy now consists of three dimensions (*political participation, political competition, freedom of opinion*) and thus closely resembles the theory of Dahl (1971). We believe that Dahl's realistic definition of democracy is an appropriate choice because it does not conceptually overlap with other institutional factors.

We compile ten regime characteristics to operationalize Dahl's seminal concept of democracy (for a summary, see Table 3). Three of them are subjective/expert-based. The other seven regime characteristics are based on objective information. Our selection thus satisfies Munck and Verkuilen's guidelines, requesting the use of both types of regime characteristics. We also give up the use of aggregated data and make sure that each regime characteristic belongs to only one of our three dimensions of democracy. Finally, we take care to ensure that our regime characteristics cover both *de facto* and *de jure* aspects.

We follow Dahl (1971) and define political participation as peoples' right to decide about their political rulers in public elections. To operationalize this dimension of democracy, we exploit three regime characteristics. The first is the share of adults with legally granted suffrage. Our primary data sources is the V-Dem database (see Coppedge et al., 2019). In addition, we collect data[23] on voter turnout and calculated the voter-population ratio to capture *de facto* restrictions of the franchise (e.g. due to material law or political repression). We admit that these two variables are imperfect measures of non-constitutional disfranchisement since a low turnout rate can also be explained by voluntary abstentions.[24] However, we also think that no other objective measures exist and that using only measures on *de jure* restrictions of the suffrage is an even worse approach.

Political competition exists when voters can choose between politicians with different party affiliations (Przeworski, 1991, Schumpeter, 1942). We operationalize this important dimension through five regime characteristics. Our first regime characteristic is a subjective index of party pluralism that distinguishes between five categories.[25] The four other regime characteristics are: (i) the share of votes not won by the strongest party/candidate,[26] (ii) the share of parliamentary seats not won by the strongest party, (iii) the share of votes won by the runner-up party/candidate divided by the share of votes won by the strongest party/candidate, and (iv) the share of seats in parliament won by the runner-up party divided by the share of seats won by the strongest party.

In line with Dahl (1971), we define the freedom of opinion as people's right to choose their sources of information and to express their political views even if these views are not compatible with the views of the government. To operationalize this third dimension of our definition, we exploit gender-specific ratings on the freedom of discussion from the V-Dem database (Coppedge et al., 2019).

### 4.4.3 Revision: priming data

In the course of our update, we also modified the procedure that we apply to identify country-year observations that are appropriate for being part of the priming data. Our selection is now based on the UDS and V-Dem's Polyarchy index instead of the Polity index. In particular, we label a regime as highly

[23] We compile our objective data from various sources. A detailed documentation can be found here: `www.ml-democracy-index.net`.

[24] Similarly, Hadenius (1992) argues that a high turnout rate does not necessarily correlate with democratic institutions. He illustrates his argument by referring to the Soviet Union in which voter turnout always exceeded 99 percent. Furthermore, since voting is compulsory in some countries, one might wonder about the appropriateness of using turnout as a regime characteristic. We share all these concerns. However, we also think that the aforementioned objections are of relatively little relevance in our case because we do not assume a linear relationship between the turnout rate and the level of democracy when applying the SVM algorithm. We rather assume a non-linear relationship and that the effect of turnout on democracy depends on the institutional environment. In Section 4.5.7, we illustrate this feature of the SVM approach.

[25] The five categories are: (i) there are no political parties, (ii) one legal party exists, (iii) there are multiple parties but opposition parties are faced with significant obstacles, (iv) there are multiple parties but opposition parties are faced with small obstacles, and (v) there are multiple parties and virtually no obstacles for opposition parties. To construct this measure, we compile information provided by the V-Dem database (Coppedge et al., 2019), the database of the Inter-Parliamentary Union (2019), and four election handbooks (see Nohlen et al., 1999, Nohlen et al., 2001, Nohlen, 2005, Nohlen and Stöver, 2010).

[26] Following Vanhanen (2000), we weight parliamentary and presidential elections according to their relevance for the political decision making process.

democratic (autocratic) if it belongs to the upper (lower) decile of either of these indices. In total, our new priming data consists of 2.975 country-year observations (for a complete list, see Table B.1).

We changed our labeling procedure for three main reasons. First, we found during some methodological and plausibility checks that using two indicators instead of one indicator slightly improves the quality of our Machine Learning indices. Second, since the Polity index has a graded scale, we cannot select a similar number of highly autocratic/democratic regimes without applying an asymmetric labeling procedure. Finally, we get a more heterogeneous sample of highly autocratic/democratic regimes when replacing the old with the new procedure.

## 4.5 Methodological discussion

### 4.5.1 Is there consensus on the regimes in the priming data?

To produce valid democracy indices, our Machine Learning method requires that the priming data includes regimes that are correctly labeled as highly democratic or highly autocratic. We proceeded in two steps to test whether our labeled country-year observations can be considered as suitably chosen. First, we checked whether our labels contradict with the descriptions in Nohlen's (1999, 2001, 2005, 2010) election handbooks, or with reports published by NGOs and international organizations. We found no country-year observation whose label is completely unreasonable and only very few cases where the label might be debatable. For example, our priming data includes Uganda in 1980 as an autocratic country-year observation. We believe that the label "autocracy" is reasonable due to military coup in May 1980, but not completely uncontroversial since a multiparty election took place in December 1980.

In our second validity test, we checked how other democracy indices evaluate the regimes included in the priming data. The list of other indicators includes the DD index, the BMR index, the Polity IV index, the FH index, and the LIED index. Our results suggest that a great consensus exist regarding the regimes in the priming data (for details, see Table B.2). For example, 99.44 percent of the democratically labeled country-year observations have a Polity score of +8 or larger.

### 4.5.2 Incorrect labels

To get a better understanding of how our method reacts if the priming includes country-year observations that are wrongly classified, we run falsification tests in which we label some democratic regimes as autocracies and some autocratic regimes as democracies. Table B.3 reports the results of one of these falsification tests. In this particular test, we falsified the labels of the 32 Irish and the 32 Cuban observations in the priming data. Columns 1 and 4 of Table B.3 show that the falsification causes, on average, only minor changes in the Machine Learning indicators. For example, the continuous index changes, on average, by only 0.010 when we replace the original priming data with the falsified priming data. The remaining columns of Table B.3 illustrate that the changes in the democracy scores are also negligible for the two countries whose labels were falsified. This robustness is notable since it implies that our Machine Learning tool identifies and automatically corrects wrongly labeled regimes, at least if their number is relatively small.

### 4.5.3 Representativeness of the priming data

Autocracies differ greatly from each other (see e.g. Cheibub et al., 2010, Geddes et al., 2014). This institutional heterogeneity is a notable challenge for us since our Machine Learning tool can only produce plausible democracy indices if the priming data includes autocratic regimes of all types. To check how well we perform in this respect, we proceeded in three steps. First, we tested whether our priming data includes civil, military, communist, and royal dictatorships as defined by Cheibub et al. (2010). Second, we checked whether we labeled both autocracies with and without elections. Third, we tested whether the autocratic regimes in the priming data differ in their durability. Table B.4 presents the results of our analysis and suggests that our labeled regimes nicely reflect the existing institutional heterogeneity among autocracies.

Institutional differences also exist among democratic regimes. We thus also checked whether our priming data includes different types of democracies. Table B.5 illustrates that this is indeed the case. In particular, we show that our democratically labeled regimes differ in their form of government (presidential, semi-presidential, parliamentarian), their total number of parliamentary chambers (unicameral, bicameral), and their voting system (proportional, majoritarian).

### 4.5.4 Alternative priming data

In our baseline version, we used the upper/lower deciles of the UDS and of V-Dem's Polyarchy index to find regimes that are highly democratic/autocratic and thus suited for being part of the priming data. We chose these indices because of their detailed scales and their great coverage rates, but also admit that the selection of the labeling criteria is to some extent arbitrary. We thus conducted various tests that analyze how our Machine Learning indicators react when we use other criteria. Table B.6 reports the results of five of these tests and suggests that our democracy indices are robust to changes in the labeling criteria. This robustness is notable but hardly surprising because of the great consensus on the regimes at the ends of the autocracy-democracy spectrum (see Section 4.5.1).

### 4.5.5 Measurement uncertainty in regime characteristics

A key problem of using objective regime characteristics is that they typically suffer from measurement error because of interpolated or inconsistent data. For example, *Adam Carr's Election Archive* indicates a turnout rate of 80.4 percent for the Argentinian presidential election in 2019, whereas the database of the *International Foundation for Electoral Systems* suggests a turnout rate of 78.9 percent. The common way of dealing with this kind of measurement error is to produce a confidence interval for each democracy score (see e.g. Treier and Jackman, 2008).

We argue that the upper and lower percentile of the distribution of indices produced by our Machine Learning approach (for details, see Section 4.3.3) can be used as bounds of a confidence interval that captures the distortions caused by the measurement errors in the raw data. To substantiate our view, we first generate randomly biased regime characteristics:

$$\widetilde{z}^j = z^j + \eta^j \qquad \text{with} \qquad \eta^j \sim \mathcal{U}\left(-\lambda \cdot \sigma_{z^j},\, +\lambda \cdot \sigma_{z^j}\right)$$

where $\lambda$ specifies the degree of measurement error and $\sigma_{z^j}$ the sample standard deviation of regime characteristic $z^j$. Afterwards, we apply our aggregation tool to the biased regime characteristics and compare the resulting Machine Learning indicators with our baseline indicators. Table B.7 presents the results of our comparison for $\lambda = 0.1$ and $\lambda = 0.2$.[27] We observe that the differences between the biased and the original Machine Learning indices are quite small and almost completely covered by our confidence intervals.

#### 4.5.6 Redundant and spurious regime characteristics

A weak point of several democracy indices is that their aggregation functions do not detect redundant regime characteristics, which in turn leads to double counting (see Munck and Verkuilen, 2002). To show how our Machine Learning approach deals with this issue, we add redundant variables to our set of regime characteristics and investigate whether such an extension has an effect on the output of our method. Columns 1 and 3 of Table B.8 suggest that this not case since we do not find notable changes in our indices if we use a variable that almost perfectly correlates with the turnout rate as an additional regime characteristic.[28] Columns 2 and 4 of Table B.8 show that our indicators also remain unchanged if we augment the set of regime characteristics by a random variable.

#### 4.5.7 Shape of the aggregation function

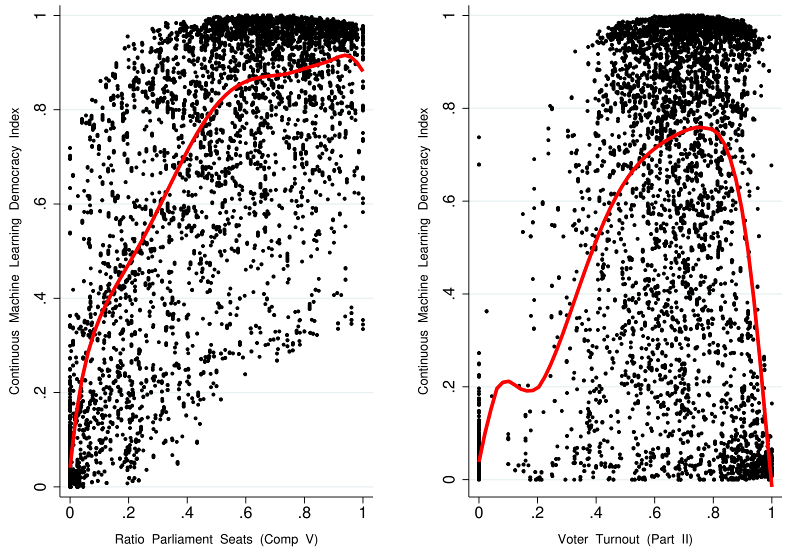
A concern about our aggregation procedure might be that we cannot provide a formula which explicitly describes how our regime characteristics affect the level of democracy, and thus cannot justify our aggregation rules through theoretical considerations. Figure 1 tries to alleviate this legitimate concern by showing scatter plots and results of bivariate non-parametric regressions. The left graph illustrates the relationship between the level of democracy (as indicated by the continuous Machine Learning index) and the regime characteristic 'Comp V'. We find that this relationship is positive and non-linear. This is plausible for two reasons: First, political competition is a central aspect of democracy. A multi-party regime should thus have a higher level of democracy than an otherwise identical single- or no-party regime. Second, small differences in the number of parliamentary seats won by a ruling and a runner-up party should not have a notable effect on the level of democracy since these differences are usually the consequence of popularity and performance differences rather than due to a systematic lack of political competition.

The right graph of Figure 1 shows that the relationship between the voter

[27] Setting $\lambda = 0.1$ implies, for example that the turnout rate can vary by $\pm 3.4$ percentage points. Due to the experiences that we made during the data collection process, we believe that this is a realistic error range.

[28] We also run this test with other regime characteristics and observe the same pattern. The summary statistics of these tests are available upon request.

**Figure 1** Functional relationship between regime characteristics and level of democracy.




**Notes**: This figure illustrates the relationship between our continuous Machine Learning index and two of our regime characteristics. The red solid lines are the results of bivariate non-parametric regressions.

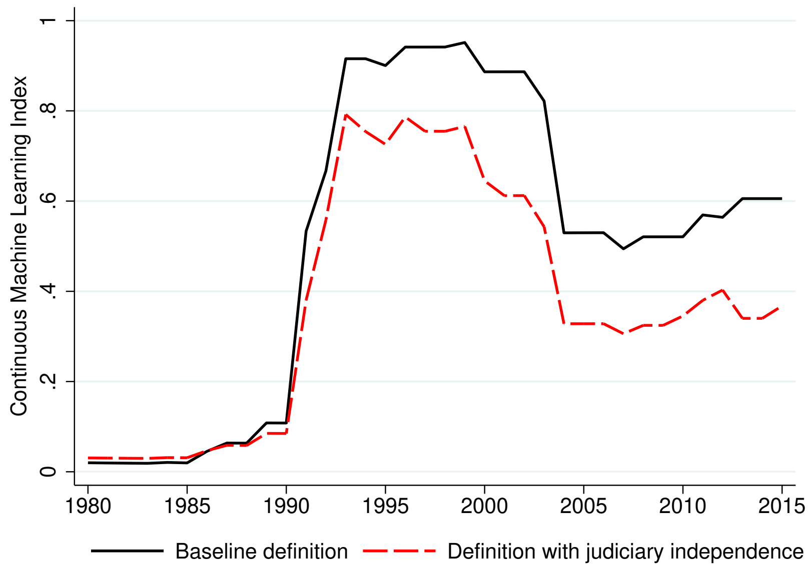
turnout and the level of democracy is non-linear as well. In particular, we see that our approach produces different democracy indicators for regimes with a very high turnout rates. This suggests that our aggregation function takes the other institutional aspects (e.g. the level of political competition) into account when deciding about how changes in political participation affect the level of democracy. We argue that such a differential treatment is reasonable because political participation is meaningless if voters cannot choose between different political parties.

### 4.5.8 Underlying concept of democracy

In Section 3, we stressed that some existing democracy indices can hardly be used in empirical investigations because of their broad concept of democracy. A potential questions is thus whether our indices suffer from the same problem. At first glance, this seems to be unlikely since we only use regime characteristics that are related to political participation, political competition, and the freedom of opinion. However, a concern might be that the democratic (autocratic) regimes that belong to the priming data do not only perform well (poorly) with regard to our three dimensions of democracy but also with regard to aspects that are typically taken into account when assuming broader concepts of democracies (e.g. judiciary independence or the freedom of press). As a consequence, it might be the case that our democracy indices unintentionally reflect a broad concept and might thus create conceptual overlaps. To check whether this problem is likely to exist, we broaden our concept of democracy by including the independence of the judiciary as a fourth dimension.[29] The rationale is that we can allay the

[29]We operationalize judiciary independence with two expert-based measures provided by V-Dem.

**Figure 2** Alternative definition of democracy.




**Notes**: This figure shows continuous Machine Learning indices for Russia. The solid black line shows the level of democracy that the SVM approach estimates when using our basic definition of democracy. The red dashed line indicates how our indices would look like if we broaden our definition of democracy by an additional aspect (judiciary independence).

aforementioned concern if we see that Machine Learning indices change due to this conceptual extension. If we do not observe any change, it is rather likely that our original indices already reflect a broad definition of democracy. Figure 2 illustrates based on the Russian indicators that our Machine Learning indices react if we modify our definition of democracy. In particular, we find that the Russian level of democracy is considerably lower when taking court independence into account.[30] We therefore believe that our original Machine Learning indices are relatively unlikely to reflect a broad concept of democracy.[31]

## 4.6 Added value of using the Machine Learning procedure

### 4.6.1 Continuous indicator

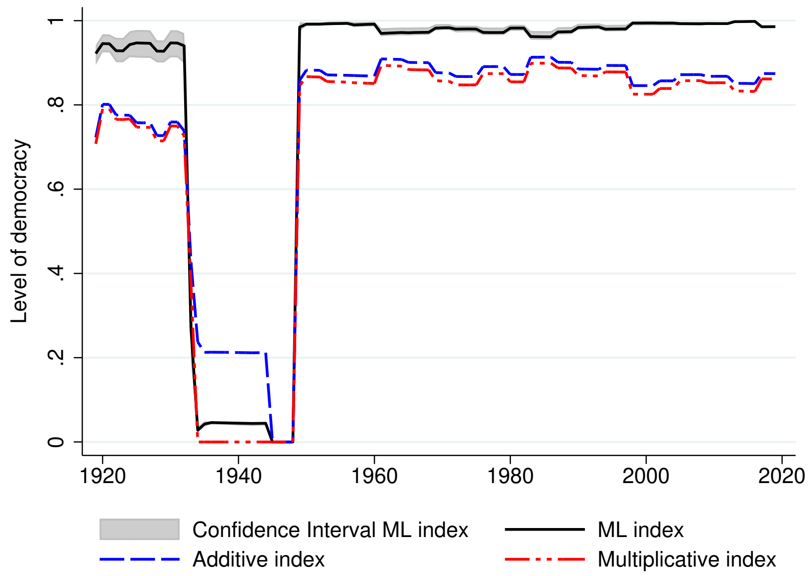
In an earlier study, we compared the performance of various data aggregation methods and showed how the choice of the aggregation technique affects the results of regression analyses (see Gründler and Krieger, 2020). A key finding of our study is that the index values for regimes at the lower/upper end of the autocracy-democracy spectrum depend crucially on the choice of the aggregation method. The explanation for these differences is that the aggregation methods differ in their assumptions about the functional relationship between the regime characteristics and the degree of democratization.

Figure 3 illustrates the difference that arises from using different aggregation

[30]We believe that this decrease is plausible because of the fairly low level of judiciary independence in Russia (Sakwa, 2010).

[31]For other countries (e.g. Hungary, Poland), we find that our conceptual change has similar consequences.

**Figure 3** Comparison of different aggregation methods (Germany)




**Notes**: This figure shows the level of democracy of Germany, depending on how we aggregate our ten regime characteristics. For further details on the construction of the additive and the multiplicative index, see Gründler and Krieger (2020).

methods based on the German level of democracy. We find that the additive measure indicates the existence of some democratic structures during the Nazi period (1933 – 1945), while the multiplicative and the Machine Learning index suggest the absence of democracy. We think that the latter assessment is more plausible because the Nazi party persecuted its opponents and heavily restricted the freedom of opinion (Shirer, 1991).[32] We also find notable difference for the Post-World-War-II period. Our Machine Learning index does not indicate a lack of democracy in Germany, whereas the additive and the multiplicative measure suggest a non-negligible lack of democracy.[33] We argue that the former result is more plausible since all elections were free and highly competitive in Germany since 1949 (Nohlen and Stöver, 2010).

In sum, we are convinced that our Machine Learning approach creates more plausible indicators for highly autocratic and highly democratic regimes since it produces a more flexible aggregation function (for more details, see Gründler and Krieger, 2020).[34] Another major advantage of our approach is that it produces confidence intervals that reflect the extent of measurement uncertainty.[35]

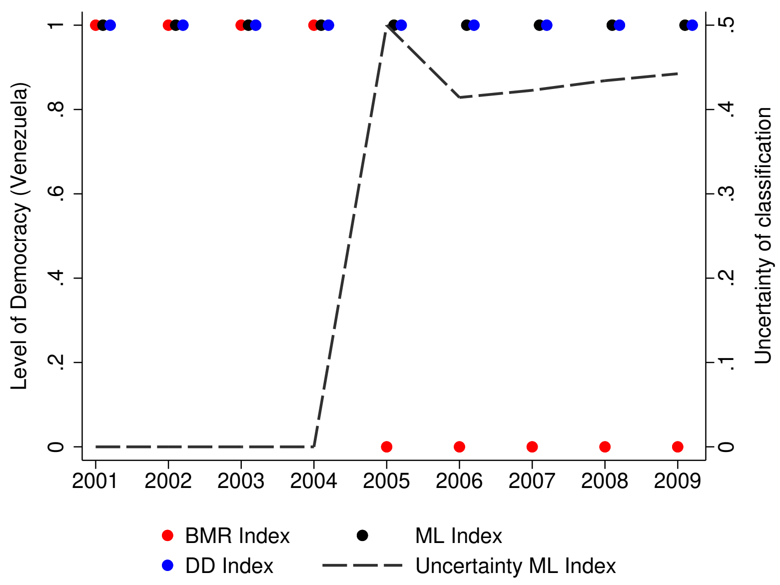
[32] The additive measure indicates a non-negligible level of democracy for Nazi Germany since the Nazi regime organized single-party elections, and because the additive approach assumes that political participation affects the level of democracy independently from the other dimensions.

[33] The additive and the multiplicative index indicate a lack of democracy since they implicitly assume that a country is only fully democratic if all regime characteristics reach their highest level. We believe that this assumption is too strict. For example, if some people voluntarily decide to abstain from voting, we can hardly argue that a lack of democracy exists.

[34] A concern might be that other aggregation tools perform better for hybrid regimes. Gründler and Krieger (2020) allay this concern by showing that all methods that have been used so far for the measurement of democracy produce rather similar indices for hybrid regimes.

[35] Figure 3 also illustrates that the level of measurement uncertainty varies across county-year observations. In particular, we observe greater measurement uncertainty for the Weimar Republic

**Figure 4** Comparison of different dichotomous indices (Venezuela, 2001 – 2009)




**Notes:** The figure illustrates how the DD index, the BMR index, and our dichotomous Machine Learning index classify Venezuela in the period from 2001 to 2009.

#### 4.6.2 Dichotomous indicator

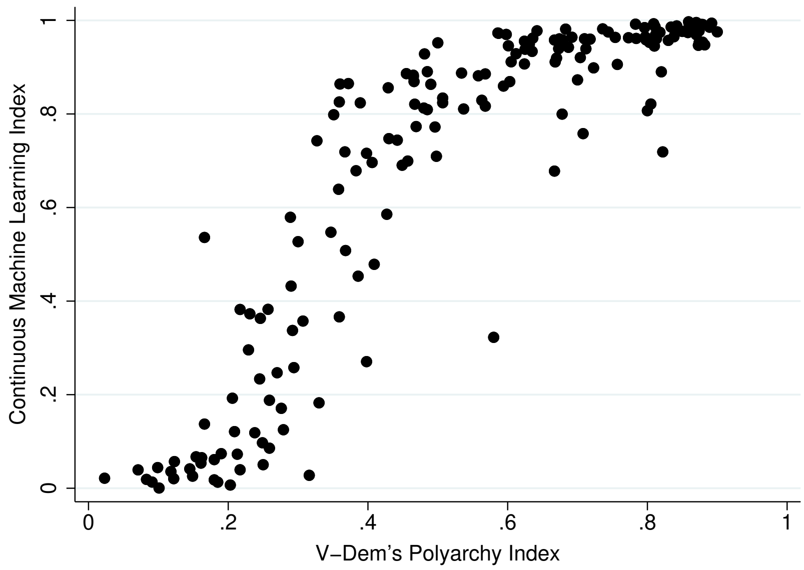
Our Machine Learning approach has three major strengths compared to other approaches that produce dichotomous indices. The first strength is that our method handles both binary and non-binary regime characteristics. The second strength is that our algorithm classifies regimes in a objective and consistent manner. The third strength is that our aggregation technique creates confidence intervals.

We believe that the third feature of our approach is the most valuable from an empirical perspective. To substantiate this view, we take a deeper look at Venezuela in 2005. At this time, Venezuela was governed by President Hugo Chávez and his socialist party. After assuming office in 1999, Hugo Chávez implemented a number of reforms that reduced the power of the Venezuelan parliament and limited the extent of political competition (Corrales, 2015, Hsieh et al., 2011, Nohlen, 2005). As a consequence, some opposition parties withdrew from the parliamentary election in 2005, which in turn further increased the influence of Hugo Chávez and his socialist party. Figure 4 illustrates that dichotomous democracy indices react differently to this event: the DD and the Machine Learning index remain unchanged, whereas the BMR index indicates a transition towards autocracy. We think that identifying the "correct" reaction is impossible since the quality of the democratic institutions deteriorated gradually in Venezuela.[36] However, we also believe that our approach has an advantage compared to the other approaches because we compute a distribution of indices for each country-year observation and can thus reflect the extent of measurement

(1919 – 1933) than for the post-World-War-II period. We think that this is plausible and consistent with the descriptions in qualitative studies.

[36]The DD (ML) index suggests that Venezuela became an autocratic regime in 2016 (2018).

**Figure 5** Comparison between the continuous Machine Learning index and V-Dem's Polyarchy index.




**Notes**: This figure compares the continuous Machine Learning index and V-Dem's Polyarchy index for the year 2019.

uncertainty. Figure 4 shows that the measurement uncertainty is high for the period from 2005 to 2009. We think that this result is plausible given that the Venezuela was a hybrid regime at this time.

### 4.7 Comparison with other indicators

To illustrate how our Machine Learning indices behave compared to established indicators, we conduct two analyses. In the first analysis, we compare our dichotomous Machine Learning index with the DD index (for details on the DD index, see Section 3.2). For 2019, we find that there is a total overlap of 184 regimes, 162 of them are equally classified (88.04%). For 9 out of the other 22 countries, the Machine Learning index indicates a relatively high degree of measurement uncertainty. This uncertainty suggests that these 9 countries are neither highly democratic nor fully autocratic and that it is thus difficult to classify them. Among the remaining countries, 4 discrepancies can be explained by the fact that the DD index classifies a regime as a democracy only if an alternation in power took place under the same election rules that brought the incumbent into power. The other discrepancies between the dichotomous Machine Learning indicator and the DD index exist because of methodological and other conceptual differences.

Our second analysis compares the continuous Machine Learning indicator with V-Dem's Polyarchy indicator. To this end, we present a scatter plot in Figure 5 that compares the two indices for 2019. We find a strong positive correlation (0.875). The most remarkable difference between the two indicators is that our Machine Learning index uses the full value range, while almost all countries are classified between 0.1 and 0.9 when considering V-Dem's Polyarchy. As outlined in great detail by Gründler and Krieger (2020), a very likely explanation for this difference is that the aggregation method applied by Teorell et al. (2019) creates implausibly low (high) values for highly democratic (autocratic) regimes (see also Section 4.6.1). From an empirical point of view, these behaviors are problematic because they cause biased estimates in OLS and 2SLS regressions (Gründler and Krieger, 2020).

### 4.8 Practitioner's guide

Our Machine Learning democracy indicators are available for 186 countries in the period from 1919 to 2019. We will update our democracy indicators on a regular basis and publish our data on our website (`www.ml-democracy-index.net`). Our website also informs about future methodological updates and presents more detailed information about the data collection process.

Our database includes a continuous and a dichotomous democracy indicator. These two indices are conceptually equivalent and can thus be used to check whether the results of an analysis depend on the scale of the democracy index. We think that such a robustness check should be done since scholars disagree about whether democracy is better measured with a continuous or with a dichotomous indicator. We do not take sides in this debate since both scaling methods have their strengths and weaknesses (for an extensive discussion, see Collier and Adcock, 1999). Practitioners should mention these pros and cons in their studies and should also provide a well-founded explanation for their choice.

Following Solt (2019), we do not only report a single indicator for each country-year observation, but also a distribution of democracy indicators. These distributions can be used to take measurement uncertainties into account. A number of approaches exist to do this. The simplest approach is to add the standard deviation of the distribution as a control variable to the regression model. A more sophisticated approach is to repeat the regression analysis with different percentiles of the distribution (see e.g. Gründler and Krieger, 2020). A third approach is to apply empirical methods that are usually used to handle multiply imputed data (see e.g. Rubin, 2004, Solt, 2019).

## 5 Conclusion

Democracy indices are a frequently used tool in the applied political economy literature. In this paper, we give an overview about nine popular measures of democracy. Our review differs from other reviews since we pay attention to aspects that are particular relevant for empirical research. We thus believe that our paper helps practitioners to make quick and well-founded decisions about which democracy indicator fits best to their purpose.

Our paper also presents a comprehensive update of the Machine Learning democracy index that we developed some years ago (see Gründler and Krieger, 2016). The new Machine Learning indicators are available for 186 countries and cover the period from 1919 to 2019. We believe that our indices nicely complement the existing measures of democracy because we can address some frequently mentioned weaknesses. For instance, our Machine Learning approach allows us to avoid simplistic assumptions about the shape of the aggregation function. Our approach is also the first that produces confidence intervals for dichotomous democracy indices.

We think that future research should try to provide more details about why democracy indicators behave differently in empirical analyses. Some first studies exist on this question, but we are convinced that there is need for more. For example, we still know relatively little about how changes in the concept of democracy affect regression results. Addressing this pending issue is of great relevance since we can thus improve our understanding of why some democratic regimes perform better than others.

## References


Abe, S. (2005). *Support Vector Machines for pattern classification.* Springer.

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American Economic Review*, 98(3):808–42.

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2015). Democracy, redistribution, and inequality. *Handbook of Income Distribution*, 2(1):1885 – 1966.

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, 127(1):47–100.

Aidt, T. S., Daunton, M., and Dutta, J. (2010). The retrenchment hypothesis and the extension of the franchise in England and Wales. *Economic Journal*, 120(547):990–1020.

Aidt, T. S. and Leon, G. (2016). The democratic window of opportunity: Evidence from riots in Sub-Saharan Africa. *Journal of Conflict Resolution*, 60(4):694–717.

Alvarez, M., Cheibub, J. A., Limongi, F., and Przeworski, A. (1996). Classifying political regimes. *Studies in Comparative International Development*, 31(2):3–36.

Arat, Z. F. (1991). *Democracy and human rights in developing countries.* Boulder and London.

Bennett, K. P. and Campbell, C. (2000). Support Vector Machines: Hype or hallelujah? *ACM SIGKDD Explorations Newsletter*, 2(2):1–13.

Bjørnskov, C. and Rode, M. (2020). Regime types and regime change: A new dataset on democracy, coups, and political institutions. *Review of International Organizations*, 15(1):531–551.

Blaug, R. and Schwarzmantel, J. J. (2016). *Democracy: A reader.* Columbia University Press.

Boese, V. A. (2019). How (not) to measure democracy. *International Area Studies Review*, 22(2):95–127.

Bogaards, M. (2012). Where to draw the line? From degree to dichotomy in measures of democracy. *Democratization*, 19(4):690–712.

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

Boix, C., Miller, M., and Rosato, S. (2018). Boix-Miller-Rosato Dichotomous Coding of Democracy, 1800-2015. *Harvard Dataverse.*

Bollen, K. A. (1999). *Cross-national indicators of liberal democracy, 1950-1990.* Inter-university Consortium for Political and Social Research.

Bollen, K. A. and Jackman, R. W. (1989). Democracy, stability, and dichotomies. *American Sociological Review*, 54(4):612–621.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory.* ACM Press.

Bowman, K., Lehoucq, F., and Mahoney, J. (2005). Measuring political democracy case expertise, data adequacy, and central america. *Comparative Political Studies*, 38(8):939–970.

Cheibub, J. A., Gandhi, J., and Vreeland, J. R. (2010). Democracy and dictatorship revisited. *Public Choice*, 143(1-2):67–101.

Cingranelli, D., Richards, D., and Clay, K. C. (2014). *The CIRI Human Rights Dataset.* CIRI Human Rights Data Project.

Collier, D. and Adcock, R. (1999). Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science*, 2:537–565.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Skaaning, S.-E., and Teorell, J. (2019). V-dem codebook v9.

Coppedge, M. and Reinicke, W. H. (1990). Measuring polyarchy. *Studies in Comparative International Development*, 25(1):51–72.

Corrales, J. (2015). The authoritarian resurgence: autocratic legalism in Venezuela. *Journal of Democracy*, 26(2):37–51.

Cortes, C. and Vapnik, V. (1995). Support-Vector networks. *Machine learning*, 20(3):273–297.

Dahl, R. A. (1971). *Polyarchy: Participation and Opposition.* Yale University Press.

De Haan, J. and Sturm, J.-E. (2003). Does more democracy lead to greater economic freedom? New evidence for developing countries. *European Journal of Political Economy*, 19(3):547–563.

Dorsch, M. T. and Maarek, P. (2019). Democratization and the conditional dynamics of income distribution. *American Political Science Review*, 113(2):385–404.

Dorsch, M. T. and Maarek, P. (2020). Economic downturns, inequality, and democratic improvements. *European Journal of Political Economy*, forthcoming.

Freedom House (2019). *Freedom in the World 2019: Methodology.*

Gasiorowski, M. J. (1996). An overview of the political regime change dataset. *Comparative Political Studies*, 29(4):469–483.

Gassebner, M., Lamla, M. J., and Vreeland, J. R. (2013). Extreme bounds of democracy. *Journal of Conflict Resolution*, 57(2):171–197.

Geddes, B., Wright, J., and Frantz, E. (2014). Autocratic breakdown and regime transitions: A new data set. *Perspectives on Politics*, 12(2):313–331.

Gibney, M., Cornett, L., Wood, R., and Haschke, P. (2013). *Political Terror Scale, 1976-2012.*

Gleditsch, K. S. and Ward, M. D. (1997). Double take a reexamination of democracy and autocracy in modern polities. *Journal of Conflict Resolution*, 41(3):361–383.

Gründler, K. and Krieger, T. (2016). Democracy and growth: Evidence from a machine learning indicator. *European Journal of Political Economy*, 45(1):85–107.

Gründler, K. and Krieger, T. (2020). Should we care (more) about data aggregation? evidence from the democracy-growth-nexus. *CESifo Working Paper Series.*

Guenther, N. and Schonlau, M. (2016). Support Vector Machines. *Stata Journal*, 16(4):917–937.

Gutmann, J. and Voigt, S. (2018). The rule of law: Measurement and deep roots. *European Journal of Political Economy*, 54(1):68–82.

Guttman, L. (1994). *Louis Guttman on theory and methodology: Selected writings.* Dartmouth Publishing Group.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using Support Vector Machines. *Machine learning*, 46(1-3):389–422.

Hadenius, A. (1992). *Democracy and development.* Cambridge University Press.

Harms, P. and Ursprung, H. W. (2002). Do civil and political repression really boost foreign direct investments? *Economic Inquiry*, 40(4):651–663.

Hsieh, C.-T., Miguel, E., Ortega, D., and Rodriguez, F. (2011). The price of political opposition: Evidence from Venezuela's maisanta. *American Economic Journal: Applied Economics*, 3(2):196–214.

Inter-Parliamentary Union (2019). Parline database on national parliaments. `http://archive.ipu.org/`.

Knutsen, C. H. (2010). Measuring effective democracy. *International Political Science Review*, 31(2):109–128.

Knutsen, C. H. and Wig, T. (2015). Government turnover and the effects of regime type: How requiring alternation in power biases against the estimated economic benefits of democracy. *Comparative Political Studies*, 48(7):882–914.

Lindberg, S. I., Coppedge, M., Gerring, J., and Teorell, J. (2014). V-dem: A new way to measure democracy. *Journal of Democracy*, 25(3):159–169.

Lührmann, A., Tannenberg, M., and Lindberg, S. I. (2018). Regimes of the world (row): Opening new avenues for the comparative study of political regimes. *Politics & Governance*, 6(1).

Márquez, X. (2018). Replicating and extending the UD scores of Pemstein, Meserve, and Melton. `https://xmarquez.github.io/QuickUDS/articles/Replicating_and_extending_the_UD_scores.html`.

Marshall, M., Gurr, T. R., and Jaggers, K. (2019). *Polity IV project. Political Regime Characteristics and Transitions, 1800-2015.* Center of Systematic Peace.

Mattera, D. and Haykin, S. (1999). Support Vector Machines for dynamic reconstruction of a chaotic system. In *Advances in kernel methods*, pages 211–241. MIT Press.

Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical transactions of the Royal Society*, 209(1):415–446.

Munck, G. L. and Verkuilen, J. (2002). Conceptualizing and meassuring democracy: Evaluating alternative indices. *Comparative Political Studies*, 35(1):5–34.

Nohlen, D. (2005). *Elections in the Americas: A data handbook.* Oxford University Press.

Nohlen, D., Grotz, F., and Hartmann, C. (2001). *Elections in Asia and the Pacific: A data handbook.* Oxford University Press.

Nohlen, D. and Stöver, P. (2010). *Elections in Europe: A data handbook.* Nomos.

Nohlen, D., Thibaut, B., and Krennerich, M. (1999). *Elections in Africa: A data handbook.* Oxford University Press.

O'Donnell, G. A. (2001). Democracy, law, and comparative politics. *Studies in Comparative International Development*, 36(1):7–36.

Pemstein, D., Meserve, S. A., and Melton, J. (2010). Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis*, 18(4):426–449.

Potrafke, N. (2012). Islam and democracy. *Public Choice*, 151(1-2):185–192.

Przeworski, A. (1991). *Democracy and the market: Political and economic reforms in Eastern Europe and Latin America.* Cambridge University Press.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Sakwa, R. (2010). The dual state in russia. *Post-Soviet Affairs*, 26(3):185–206.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy.* Harper & Brothers.

Shirer, W. L. (1991). *The rise and fall of the Third Reich: A history of Nazi Germany.*

Random House.

Skaaning, S.-E., Gerring, J., and Bartusevičius, H. (2015). A lexical index of electoral democracy. *Comparative Political Studies*, 48(12):1491–1525.

Solt, F. (2019). Measuring income inequality across countries and over time: The standardized world income inequality database.

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer.

Teorell, J., Coppedge, M., Lindberg, S., and Skaaning, S.-E. (2019). Measuring polyarchy across the globe, 1900–2017. *Studies in Comparative International Development*, 54(1):71–95.

Treier, S. and Jackman, S. (2008). Democracy as a latent variable. *American Journal of Political Science*, 52(1):201–217.

Vanhanen, T. (2000). A new dataset for measuring democracy, 1810-1998. *Journal of Peace Research*, 37(2):251–265.

Vanhanen, T. (2019). *Measures of Democracy 1810-2018, version 8.0.* Finish Social Science Data Archive, Tampere (FI).

Vapnik, V. (1995). *The nature of statical learning theory*. Springer.

Vapnik, V. (1998). *Statistical learning theory*. Wiley.

Vapnik, V. and Lerner, A. (1963). Generalized Portrait Method for pattern recognition. *Automation and Remote Control*, 24(6):774–780.

Zuazu, I. (2019). The growth effect of democracy and technology: An industry disaggregated approach. *European Journal of Political Economy*, 56:115–131.

# Appendix for online publication

## A Support Vector Machines

In this supplementary section, we provide a short formal introduction to Support Vector Classifications and Support Vector Regressions. Our description is based on the textbooks by Abe (2005), Steinwart and Christmann (2008), and Vapnik (1995, 1998).[37]

### A.1 Support Vector Classifications

The Support Vector Classification is a non-linear extension of the General Portrait Algorithm (GPA). The basic GPA was proposed by Vapnik and Lerner (1963) and assumes the existence of a hyperplane

$$E_{\mathbf{w},b}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \qquad \mathbf{w} \in \mathbb{R}^m \,,\ ||\mathbf{w}|| = 1 \,,\ b \in \mathbb{R} \,,\ \mathbf{x} \in \mathbb{R}^m \tag{2}$$

that separates the observations in sample $\mathcal{S} = \{(\mathbf{x}_1, y_1) \ldots (\mathbf{x}_n, y_n)\}$ according to their labels $y \in \{-1, 1\}$.[38] Graph I in Figure A.1 shows a unidimensional case in which such a separation is possible.

The primary objective of the GPA is to find a linear classification function that assigns any input $\mathbf{x}_i$ to its output $y_i$ for all $n > 1$ observations in the sample $\mathcal{S}$. However, as Graph II in Figure A.1 indicates, the number of eligible hyperplanes might be infinite. To find a unique solution, the GPA first computes the distance (*margin*) between each separating hyperplane and its nearest observation, and then chooses the hyperplane with the greatest margin in $\mathcal{S}$ (see Graphs III and IV in Figure A.1). Formally, this means that the GPA solves the quadratic optimization problem:

$$\min_{\mathbf{w},\, b} \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle \qquad \text{s.t.} \qquad y_i \cdot (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 \tag{3}$$

and that it uses the solution $(\mathbf{w}^*, b^*)$ to calculate the classification function:

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}(\langle \mathbf{w}^*, \mathbf{x} \rangle + b^*) \qquad \text{where} \qquad \mathbf{w}^* \in \mathbb{R}^m \quad \text{and} \quad b^* \in \mathbb{R}. \tag{4}$$

In applied research fields, the GPA attracts only little attention since a linear separation does not exist in general (see Figure A.2, Graph I). Boser et al. (1992) therefore enhanced the original GPA such that it computes non-linear classification functions. The idea behind this extension is simple: Boser et al. (1992) first use a non-linear function $\Phi \colon \mathbb{R}^m \to \mathcal{H}$ to map the input characteristics $\mathbf{x} \in \mathbb{R}^m$ onto a *Reproducing Hilbert Space* $(\mathcal{H})$[39] and then apply the GPA to the adjusted sample $\mathcal{S}_{\mathcal{H}} = \{(\Phi(\mathbf{x}_i), z_i) \,|\, i = 1, \ldots, n\}$ to get a non-linear classification function (for a graphical illustration, see Graphs II & III of Figure A.2).

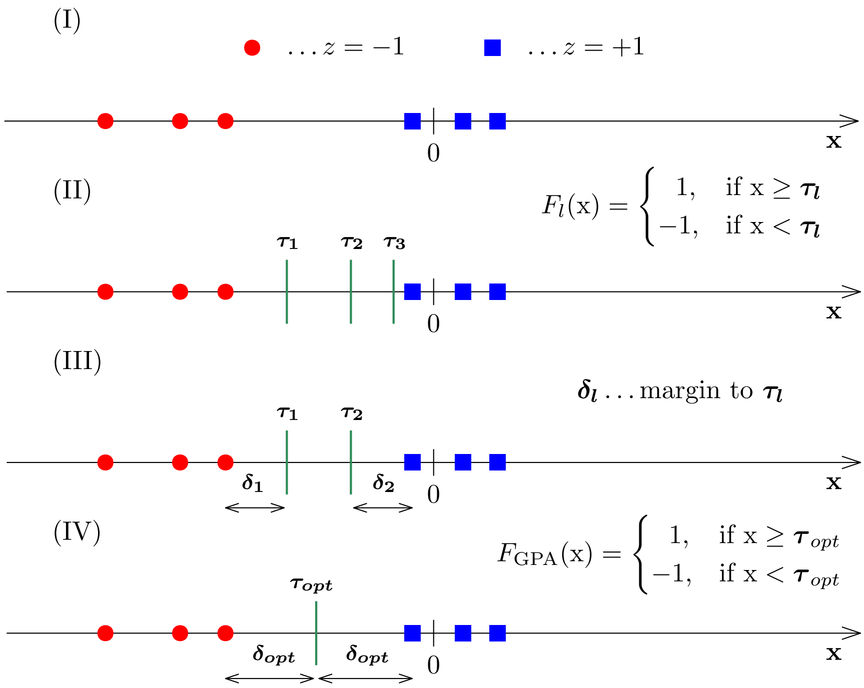
Cortes and Vapnik (1995) suggest that random noise and measurement error may lead to mislabeling. They therefore relax the auxiliary conditions of the GPA by

[37] Note that we already presented a similar introduction to SVM in one of our related papers (see Gründler and Krieger, 2020).

[38] Note that $\langle \cdot, \cdot \rangle$ indicates the dot product of two vectors.

[39] The non-linear extension suggested by Boser et al. (1992) is based on mathematical theorems that prove the existence of a *feature space* $\mathcal{H}$, in which a hyperplane can perfectly separate the sample data $\mathcal{S}$. For details, see Steinwart and Christmann (2008).

**Figure A.1** Linear separation — One-dimensional case.




**Notes:** Graph I is a one-dimensional example in which the GPA is applicable. Graph II shows that more than one hyperplane may separate the observations according to their labels. Graph III explains how the margin $\delta$ is calculated. Graph IV illustrates that the GPA selects the hyperplane with the largest margin.

including slack variables $\xi_i \geq 0$. Together with the non-linear GPA extension of Boser et al. (1992), this adjustment yields the optimization problem:

$$\min_{\mathbf{w}_{\mathcal{H}}, b_{\mathcal{H}}, \xi} \frac{1}{2}\langle \mathbf{w}_{\mathcal{H}}, \mathbf{w}_{\mathcal{H}} \rangle + C \cdot \sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad y_i \cdot (\langle \mathbf{w}_{\mathcal{H}}, \Phi(\mathbf{x}_i) \rangle + b_{\mathcal{H}}) \geq 1 - \xi_i \quad \forall\, i \tag{5}$$

where $C$ denotes a fixed cost parameter for penalizing misclassifications.

When the dimension of $\mathcal{H}$ is very large, solving (5) might be computationally infeasible. The standard solution for this problem is to consider the corresponding dual program:

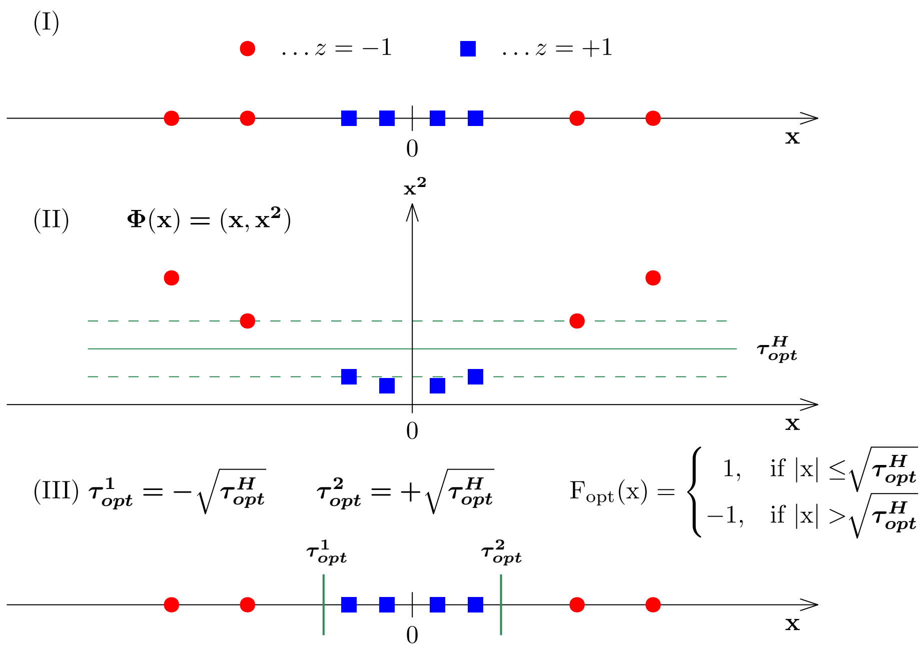
$$\max_{\alpha \in [0,C]^n} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \cdot \sum_{i,j=1}^{n} z_i \cdot z_j \cdot \alpha_i \cdot \alpha_j \cdot \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \quad \text{s.t.} \quad \sum_{i=1}^{n} z_i \cdot \alpha_i = 0 \tag{6}$$

where $\alpha_1, \dots, \alpha_n$ are the Lagrange multipliers of the primal program. The closed form solution of the dual program for the classification function is:

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}\left( \sum_{i=1}^{n} z_i \alpha_i^* \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}} + b_{\mathcal{H}}^* \right). \tag{7}$$

The problem when computing (7) is that an appropriate *feature map* $\Phi\colon \mathbb{R}^m \to \mathcal{H}$ is usually not known. Schölkopf et al. (1998) therefore replace the unknown inner

**Figure A.2** Non-linear separation — One-dimensional case.




**Notes:** Graph I shows an example in which the GPA is not applicable in $\mathcal{X} = \mathbb{R}$. In Graph II, a function $\Phi(x) = (x, x^2)$ is used to map the input data from $\mathcal{X} = \mathbb{R}$ onto a feature space $\mathcal{H} = \mathbb{R}^2$ and GPA computes a dividing hyperplane in $\mathcal{H}$. Graph III illustrates that the linear solution in $\mathcal{H}$ implies a non-linear solution in $\mathcal{X}$.

product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}}$ with a known *kernel function* $\mathfrak{K}\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$:

$$\mathfrak{F}(\mathbf{x}) \;=\; \text{sign}\left( \sum_{i=1}^{n} z_i \cdot \alpha_i^* \cdot \mathfrak{K}(\mathbf{x}_i, \mathbf{x}) \;+\; b_{\mathcal{H}}^* \right).^{40} \tag{8}$$

An observation is called *Support Vector* if its Lagrange multiplier $\alpha_i^*$ is nonzero. The algorithm takes its name from these observations because only Support Vectors influence the shape of the classification function.

## A.2 Support Vector Regressions

In its basic form, the GPA is limited to applications in which the output variable comes from a countably finite set. Vapnik (1995, 1998) addresses this problem by introducing a variant of the GPA that estimates real-valued functions. The key objective of this procedure is to find a function $\mathfrak{F}\colon \mathcal{X} \subseteq \mathbb{R}^m \to \mathcal{Y} \subseteq \mathbb{R}$ whose predicted outcomes deviate at most by $\varepsilon \geq 0$ from the true outcome for all observations in the sample $\mathcal{S} = \{(\mathbf{x}_1, y_1) \ldots (\mathbf{x}_n, y_n)\}$:

$$|\,\mathfrak{F}(\mathbf{x}_i) \,-\, y_i\,| \;\overset{!}{\leq}\; \varepsilon \qquad \forall\, i = 1, \ldots, n. \tag{9}$$

[40] The idea by Schölkopf et al. (1998), known as *kernel trick*, is based on a paper of Mercer (1909) who found out that each kernel function $\mathfrak{K}\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is related to a Reproducing Hilbert Space $\mathcal{H}$ with:

$$\mathfrak{K}(\mathbf{x}_i, \mathbf{x}_j) \;=\; \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \quad \forall\, \mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^m.$$

Achieving this goal is simple if the regression function is a hyperplane:

$$\mathfrak{F}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \qquad \text{with} \qquad \mathbf{w} \in \mathbb{R}^m \quad \text{and} \quad b \in \mathbb{R} \tag{10}$$

since one only has to compute the solution $(\mathbf{w}^*, b^*)$ of the quadratic optimization problem:

$$\min_{\mathbf{w}, b} \frac{1}{2} \cdot ||\, w \,||^2 \qquad \text{s.t.} \qquad \begin{cases} z_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon & \forall\, i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - z_i \leq \varepsilon & \forall\, i. \end{cases} \tag{11}$$

However, this approach has the problem that solving (11) often turns out to be impossible in practical applications. Vapnik (1995, 1998) therefore introduces slack variables $(\xi_i^+, \xi_i^-) \in \mathbb{R}_+^2$ that relax the auxiliary conditions and uses a feature map $\Phi\colon \mathcal{X} \to \mathcal{H}$ that allows for non-linear estimations. These extension change the optimization problem to:

$$\min_{\mathbf{w}, b, \xi_i^+, \xi_i^-} \frac{1}{2} ||\, w \,||^2 + C \cdot \sum_{i=1}^{n} \left( \xi_i^+ + \xi_i^- \right) \quad \text{s.t.} \quad \begin{cases} z_i - \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^+ \\ \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b - z_i \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0. \end{cases} \tag{12}$$

The corresponding dual problem is

$$\max_{\alpha^+, \alpha^-} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-) \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}} - \varepsilon \sum_{i=1}^{n} (\alpha_i^+ + \alpha_i^-) + \sum_{i=1}^{n} y_i (\alpha_i^+ - \alpha_i^-)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} (\alpha_i^+ - \alpha_i^-) = 0 \quad \text{and} \quad \alpha_i^+, \alpha_i^- \in [0, C],$$

where we denote by $\alpha^+ = (\alpha_1^+, \ldots, \alpha_n^+)$ and $\alpha^- = (\alpha_1^-, \ldots, \alpha_n^-)$ the Lagrangian multipliers of the primal program. The closed form solution of the dual program is:

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n} \left( \alpha_i^+ - \alpha_i^- \right) \cdot \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}} + b_{\mathcal{H}}^*. \tag{13}$$

Since the mapping function $\Phi\colon \mathcal{X} \to \mathcal{H}$ is not known, Vapnik (1995, 1998) replaces the unknown inner product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}}$ with a kernel $\mathfrak{K}\colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$. The shape of the non-linear regression function

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \cdot \mathfrak{K}(\mathbf{x}_i, \mathbf{x}) + b_{\mathcal{H}}^*. \tag{14}$$

is only depending on those observations (called *Support Vectors*) whose Lagrangian multipliers $(\alpha_i, \alpha_i^*)$ are different from zero.

# B Additional tables

**Table B.1** Priming data — Selected country-years.

| Country | Observations | Years |
|---|---|---|
| | *Democratic regimes* | |
| Australia | 53 | 1950, 1962 – 1964, 1967 –1968, 1972 – 2018 |
| Austria | 68 | 1950 – 2017 |
| Bahamas | 1 | 2017 |
| Barbados | 1 | 2017 |
| Belgium | 70 | 1950 – 2019 |
| Brazil | 24 | 1991, 1993 – 2015 |
| Canada | 27 | 1976 – 1979, 1988 – 1992, 2002 – 2019 |
| Chile | 19 | 1998, 2000 – 2017 |
| Costa Rica | 40 | 1980 – 2019 |
| Cyprus | 7 | 2008 – 2009, 2011 – 2013, 2015, 2017 |
| Czech Republic | 23 | 1991 – 2012, 2014 |
| Denmark | 70 | 1950 – 2019 |
| Dominica | 1 | 2017 |
| Estonia | 26 | 1993 – 2012, 2014 – 2019 |
| Finland | 50 | 1967 – 1971, 1975 – 2019 |
| France | 30 | 1975 – 1977, 1988 – 1996, 2002 – 2019 |
| Germany | 61 | 1957 – 2017 |
| Greece | 34 | 1986 – 2019 |
| Hungary | 10 | 1994 – 1997, 2002 – 2005, 2007 – 2008 |
| Iceland | 48 | 1971 – 2018 |
| Ireland | 32 | 1987 – 1988, 1990 – 2019 |
| Italy | 65 | 1950 – 2007, 2013 – 2019 |
| Japan | 28 | 1959 – 1960, 1967 – 1969, 1973 – 1976, 1980 – 1982, 1990 – 1992, 1997 – 2004, 2010 – 2013, 2017 |
| Korea (South) | 8 | 1998 – 2000, 2004 – 2007, 2018 |
| Latvia | 3 | 2013, 2015 – 2016 |
| Luxembourg | 46 | 1950 – 1953, 1955, 1960 – 1963, 1974 – 1979, 1984 – 1993, 1999 – 2019 |
| Malta | 13 | 2000, 2003 – 2008, 2010 – 2011, 2014 – 2017 |
| Netherlands | 63 | 1956 – 2018 |
| New Zealand | 46 | 1950, 1973, 1975 – 2017, 2019 |
| Norway | 60 | 1960 – 2019 |
| Poland | 22 | 1992 – 1997, 2000 – 2015 |
| Portugal | 38 | 1982 – 2019 |
| Slovakia | 7 | 2004 – 2006, 2008, 2010 – 2011, 2016 |
| Slovenia | 20 | 1992 – 2000, 2002 – 2004, 2009 – 2016 |
| Spain | 40 | 1980 – 2019 |
| St. Kitts and Nevis | 7 | 2005, 2010 – 2014, 2017 |
| St. Lucia | 6 | 2009 – 2010, 2013 – 2015, 2017 |
| St. Vincent and the Grenadines | 3 | 2013 – 2014, 2017 |
| Sweden | 70 | 1950 – 2019 |
| Switzerland | 39 | 1980 – 2000, 2002 – 2019 |
| United Kingdom | 39 | 1950 – 1954, 1961, 1964, 1979, 1987 – 1996, 1999 – 2019 |
| United States | 30 | 1983 – 1992, 1994 – 2000, 2004 – 2016 |
| Uruguay | 30 | 1989 – 2017, 2019 |
| | *Autocratic regimes* | |
| Afghanistan | 38 | 1950 – 1963, 1978 – 2001 |
| Albania | 40 | 1950 – 1989 |
| Algeria | 19 | 1965 – 1976, 1980 – 1981, 1983 – 1987 |
| Angola | 18 | 1975 – 1992 |
| Argentina | 13 | 1966 – 1972, 1977 – 1982 |
| ⋮ | ⋮ | ⋮ |

**Table B.1** Priming data — Selected country-years (continued).

| Country | Observations | Years |
|---|---|---|
| ⋮ | ⋮ | ⋮ |
| Bahrain | 29 | 1971 – 1972, 1975 – 2000, 2016 |
| Benin | 7 | 1973 – 1979 |
| Bhutan | 54 | 1950 – 2003 |
| Bolivia | 6 | 1972 – 1977 |
| Brazil | 4 | 1965 – 1966, 1968 – 1969 |
| Brunei | 14 | 1984, 1986 – 1995, 1997, 1999, 2001 |
| Burma (Myanmar) | 36 | 1963 – 1974, 1983 – 1987, 1989, 1991 – 2008 |
| Burundi | 19 | 1967 – 1981, 1988 – 1991 |
| Cambodia | 20 | 1953, 1957 – 1969, 1975, 1980, 1982 – 1985 |
| Central African Rep. | 15 | 1966 – 1979, 2017 |
| Chad | 28 | 1962 – 1974, 1976 – 1989, 2017 |
| Chile | 14 | 1974 – 1987 |
| China | 47 | 1950 – 1980, 1990 – 1991, 2000 – 2001, 2003 – 2004, 2009, 2011 – 2019 |
| Democratic Rep. of Congo | 26 | 1965 – 1989, 2017 |
| Rep. of Congo | 5 | 1969 – 1972, 1977 |
| Cuba | 32 | 1961 – 1992 |
| Dominican Republic | 11 | 1950 – 1960 |
| Egypt | 4 | 1953 – 1956 |
| Equatorial Guinea | 11 | 1973 – 1982, 2017 |
| Eritrea | 18 | 2002 – 2019 |
| Ethiopia | 37 | 1950 – 1986 |
| Gabon | 10 | 1968 – 1972, 1981, 1983 – 1984, 1986 – 1987 |
| Germany (East) | 29 | 1960 – 1988 |
| Ghana | 1 | 1965 |
| Greece | 6 | 1968 – 1973 |
| Grenada | 4 | 1980 – 1983 |
| Guatemala | 4 | 1956, 1964 – 1965, 1983 |
| Guinea | 27 | 1958 – 1984 |
| Haiti | 26 | 1950, 1961 – 1985 |
| Iran | 25 | 1954 – 1978 |
| Iraq | 40 | 1963 – 2002 |
| Ivory Coast | 25 | 1960 – 1974, 1980 – 1989 |
| Jordan | 33 | 1950, 1957 – 1988 |
| Korea (North) | 58 | 1958 – 1962, 1966 – 2018 |
| Korea (South) | 1 | 1972 |
| Kuwait | 19 | 1961 – 1962, 1965 – 1970, 1976 – 1980, 1986 – 1991 |
| Laos | 39 | 1976 – 1988, 1991 – 2016 |
| Lesotho | 3 | 1970 – 1972 |
| Liberia | 3 | 1981 – 1983 |
| Lybia | 47 | 1951 – 1955, 1969 – 2010 |
| Malawi | 29 | 1964 – 1992 |
| Maldives | 2 | 1965 – 1966 |
| Mauritania | 6 | 1979 – 1984 |
| Mongolia | 2 | 1950 – 1951 |
| Morocco | 19 | 1956 – 1962, 1965 – 1976 |
| Mozambique | 10 | 1975 – 1980, 1982 – 1985 |
| Nepal | 22 | 1950 – 1951, 1960 – 1979 |
| Niger | 5 | 1976, 1982 – 1985 |
| Oman | 32 | 1970 – 2001 |
| Panama | 3 | 1969 – 1971 |
| Paraguay | 7 | 1954 – 1960 |
| Peru | 2 | 1969, 1973 |
| Philippines | 6 | 1972 – 1977 |
| Portugal | 20 | 1950 – 1959, 1961 – 1964, 1966 – 1971 |
| Qatar | 49 | 1971 – 2019 |
| Russia | 10 | 1950 – 1953, 1963, 1965 – 1969 |
| Rwanda | 5 | 1974 – 1978 |
| ⋮ | ⋮ | ⋮ |

**Table B.1** Priming data — Selected country-years (continued).

| Country | Observations | Years |
|---|---|---|
| ⋮ | ⋮ | ⋮ |
| Sao Tome and Principe | 6 | 1983 – 1988 |
| Saudi Arabia | 70 | 1950 – 2019 |
| Serbia | 14 | 1950 – 1952, 1954 – 1957, 1964 – 1966, 1968 – 1969, 1974, 1978 |
| Somalia | 13 | 1970 – 1979, 1986, 2011 2017 |
| South Sudan | 1 | 2017 |
| Spain | 18 | 1950 – 1967 |
| Sudan | 13 | 1959 – 1964, 1990 – 1995, 2017 |
| Swaziland | 31 | 1973 – 1999, 2001 – 2003, 2005 |
| Syria | 34 | 1961, 1970 – 1975, 1978 – 1999, 2013 – 2017 |
| Taiwan | 20 | 1950 – 1969 |
| Tajikistan | 1 | 2017 |
| Togo | 12 | 1968 – 1979 |
| Tonga | 1 | 1970 |
| Tunisia | 15 | 1956 – 1959, 1961 – 1964, 1966 – 1969, 1971, 1979 – 1980 |
| Turkmenistan | 25 | 1992 – 2013, 2015 – 2017 |
| Uganda | 9 | 1972 – 1980 |
| United Arab Emirates | 45 | 1971 – 2015 |
| Uruguay | 4 | 1976 – 1979 |
| Uzbekistan | 25 | 1993 – 2017 |
| Venezuela | 5 | 1953 – 1957 |
| Vietnam | 6 | 1954 – 1959 |
| Yemen | 33 | 1950 – 1970, 1977 – 1987, 2017 |
| Zambia | 2 | 1981, 1989 |

**Notes**: This table reports the country-year observations that are part of the priming data. The selection is based on the UDS and V-Dem's Polyarchy index.

**Table B.2** Agreement among democracy indices about labeled observations.

| | **Democratic regimes** | | | | |
|---|---|---|---|---|---|
| | Polity ($\geq 7$) | LIED ($= 6$) | FH (*Free*) | BMR ($= 1$) | BR ($= 1$) |
| **Overlap** | 0.9944 | 1.0000 | 0.9924 | 1.0000 | 1.0000 |
| | **Autocratic regimes** | | | | |
| | Polity ($\leq -7$) | LIED ($\leq 1$) | FH (*Not Free*) | BMR ($= 0$) | BR ($= 0$) |
| **Overlap** | 0.9511 | 0.9244 | 0.8976 | 0.9994 | 1.0000 |

**Notes**: This table shows agreement rates, i.e. the share of country-year observations in the priming data that are labeled as "autocratic" ("democratic") and are classified as autocracy (democracy) by an alternative index. The list of alternative indices includes: the Polity IV index, the Lexical Index of Electoral Democracy (Skaaning et al., 2015), the Freedom House index, and the dichotomous indices by Boix et al. (2013), and Bjørnskov and Rode (2020).

**Table B.3** Mislabeled regimes in priming data

| | **Continuous Indicator** | | | **Dichotomous Indicator** | | |
|---|---|---|---|---|---|---|
| | *Total* | *Cuba* | *Ireland* | *Total* | *Cuba* | *Ireland* |
| **Mean Abs. Dev.** | 0.003 | 0.002 | 0.007 | 0.017 | 0.000 | 0.000 |
| **Coverage rate** | 0.994 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Notes**: This table shows how our Machine Learning indicators reacts if the priming data includes wrongly classified regimes. To illustrate this reaction, we falsified all Irish and Cuban observations in the priming data.

**Table B.4** Representativeness of autocratically labeled regimes.

| | **Regime** | **Elections** | | **Duration** | | |
|---|---|---|---|---|---|---|
| | $\Sigma$ | *Yes* | *No* | $\leq$ *5 years* | *6 – 25 years* | $>$ *25 years* |
| **Civil dictator.** | 311 | 242 | 69 | 92 | 190 | 29 |
| **Communist dictator.** | 365 | 241 | 124 | 50 | 176 | 139 |
| **Military dictator.** | 420 | 116 | 304 | 162 | 210 | 48 |
| **Royal dictator.** | 501 | 140 | 361 | 43 | 173 | 285 |
| | *1,597* | *739* | *858* | *347* | *749* | *501* |

**Notes**: Our Machine Learning approach requires that the priming data reflects the institutional heterogeneity among autocracies and democracies. This table shows how the autocratically labeled country-year observations are distributed over different categories. The data comes from Bjørnskov and Rode (2020) and Geddes et al. (2014).

**Table B.5** Representativeness of democratically labeled regimes.

| | **Regime** | **Legislature** | | **Voting system** | |
|---|---|---|---|---|---|
| | $\Sigma$ | *Unicameral* | *Bicameral* | *Proportional* | *Majoritarian* |
| **Parliamentary** | 876 | 367 | 509 | 675 | 201 |
| **Semi-presidential** | 305 | 120 | 185 | 275 | 30 |
| **Presidential** | 197 | 55 | 142 | 159 | 38 |
| | *1,378* | *542* | *836* | *1,109* | *269* |

**Notes**: Our Machine Learning approach requires that the priming data reflects the institutional heterogeneity among autocracies and democracies. This table shows how the democratically labeled country-year observations are distributed over different categories. The data comes from Bjørnskov and Rode (2020).

**Table B.6** Alternative labeling criteria

| | UDS | Polyarchy | Add Polity | 7.5/92.5 | 12.5/87.5 |
|---|---|---|---|---|---|
| | *Continuous Machine Learning indicator* | | | | |
| Mean Abs. Dev. | 0.009 | 0.012 | 0.012 | 0.017 | 0.007 |
| Coverage rate | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | *Dichotomous Machine Learning indicator* | | | | |
| Mean Abs. Dev. | 0.006 | 0.006 | 0.012 | 0.015 | 0.003 |
| Coverage rate | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Notes**: This table shows how our Machine Learning indicators reacts if we adjust the criterion that we use to specify the priming data. In our basic specification, we use the upper and lower deciles of the UDS index and V-Dem's Polyarchy index. In Columns 1 and 2, we use only one of these indices. In Column 3, we use the Polity index as additional index. In Columns 4 (5), we label the upper and lower 7.5 (12.5) percent of the distributions.

**Table B.7** Measurement error in raw data

| | **Continuous Indicator** | | **Dichotomous Indicator** | |
|---|---|---|---|---|
| | $\lambda = 0.1$ | $\lambda = 0.2$ | $\lambda = 0.1$ | $\lambda = 0.2$ |
| **Mean Abs. Dev.** | 0.006 | 0.012 | 0.004 | 0.007 |
| **Coverage rate** | 0.996 | 0.974 | 1.000 | 1.000 |

**Notes**: This table shows how our Machine Learning indicators reacts if the regime characteristics suffer from measurement error.

**Table B.8** Random and redundant regime characteristics.

| | **Continuous Indicator** | | **Dichotomous Indicator** | |
|---|---|---|---|---|
| | *Redundant* | *Random* | *Redundant* | *Random* |
| **Mean Abs. Dev.** | 0.007 | 0.004 | 0.009 | 0.0038 |
| **Coverage rate** | 0.992 | 1.000 | 1.000 | 1.000 |

**Notes**: This table shows how our Machine Learning indicators reacts if augment our set of regime characteristics by a redundant/spurious variable.

**Table B.9** Frequently used measures of democracy

| Name | Provider | Time span | Definition | Scale | Aggregation rule | Conceptual overlaps | Confidence intervals |
|---|---|---|---|---|---|---|---|
| Boix-Miller-Rosato index | Boix et al. (2018) | 1800 – 2015 | Narrow | Dichotomous | Multiplicative approach | No | No |
| Democracy-Dictatorship index | Bjørnskov and Rode (2020) | 1950 – 2018 | Narrow | Dichotomous | Multiplicative approach | No | No |
| Polity IV index | Marshall et al. (2019) | 1800 – 2018 | Broad | Quasi-continuous | Additive approach | Yes | No |
| Freedom House indicators | Freedom House (2019) | 1972 – 2020 | Broad | Quasi-continuous/ordinal | Additive approach | Yes | No |
| Acemoglu-Naidu-Restrepo-Robinson index | Acemoglu et al. (2019) | 1960 – 2010 | Broad (inconsistent) | Dichotomous | Heuristic that combines different indices | Yes | No |
| Vanhanen index | Vanhanen (2019) | 1810 – 2018 | Narrow | Continuous | Multiplicative approach | No | No |
| Unified Democracy Score | Márquez (2018) | 1815 – 2018 | Broad (inconsistent) | Continuous | Bayesian latent variable approach | Yes | Yes |
| Lexical Index of Electoral Democracy | Skaaning et al. (2015) | 1789 – 2019 | Realistic | Ordinal | Hierarchical approach | No | No |
| V-Dem's Polyarchy index | Coppedge et al. (2019) | 1789 – 2019 | Realistic | Continuous | Combination of additive and multiplicative approach | No | Yes |

**Notes**: This table gives a brief overview about the nine most popular measures of democracy. In particular, we indicate who publishes the last update of this index and indicate the time span of each index. We also provide information about the conceptual and methodological aspects of each index. For a more extensive description, see Section 3.

// IMPRINT

**ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim**

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de