CESifo WORKING PAPERS

7480
2019

Original Version: January 2019
This Version: March 2020

Should We Care (More) About Data Aggregation? Evidence from Democracy Indices.

Klaus Gründler, Tommy Krieger

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7480*
Category 12: Empirical and Theoretical Methods

Should We Care (More) About Data Aggregation? Evidence from the Democracy Indices.

Abstract

Aggregation tools transform multidimensional data into indices. To investigate how the design of an aggregation process influences regression results, we build democracy indices that differ regarding their scale and aggregation function. Our results show that the shape of the aggregation function significantly affects OLS and 2SLS estimates since different shapes produce systematically different index values for regimes at the lower and upper end of the democracy spectrum. We also find that dichotomous indices produce significantly smaller OLS estimates than continuous indices because of greater measurement uncertainty. Whether continuous and dichotomous indices cause different 2SLS estimates depends on their design.

JEL-Codes: C260, C430, O100, O430, P160, P480.

Keywords: aggregation, data transformation, democracy, economic development, indices, machine learning, measurement of democracy, political transitions, scaling.

Klaus Gründler
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstrasse 5
Germany – 81679 Munich
gruendler@ifo.de

Tommy Krieger
ZEW Mannheim
Department of Corporate Taxation and Public Finance, L7 1
Germany – 68161 Mannheim
tommy.krieger@zew.de

March 27, 2020
We are very grateful for the constructive comments from Toke Aidt, Norbert Berthold, Richard Bluhm, Roger Congelton, Martin Gassebner, Kai Gehring, Jerg Gutmann, Raphael Franck, Stephan Maurer, Niklas Potrafke, Guido Schwerdt, Uwe Sunde, and Heinrich Ursprung. We would also like to thank seminar participants at the Universities of Cambridge, Friedrichshafen, Mannheim, Munich, and Konstanz, as well as various conference participants.

1 Introduction

Any empirical study that investiages the causes and the consequences of political transitions from autocracy to democracy (and vice versa) requires an index that measures institutional change. Creating such an index is difficult for many reasons (Coppedge et al., 2011, Munck and Verkuilen, 2002). A key challenge is to find an aggregation procedure that transforms the observable regime characteristics into an uni-dimensional index. We show how decisions made during the data aggregation process affect the results of regression analyses in which democracy indices serve as explanatory variables.

Data aggregation requires two decisions (Hawken and Munck, 2013, Munck and Verkuilen, 2002): the first decision concerns the numerical form of the indicator, while the second decision relates to the shape of the function that determines the relationship between the regime characteristics and the level of democracy. Our findings suggest that both choices have significant consequences for the results of OLS and 2SLS regressions. We also show why these decisions matter and provide some general advice regarding the judicious use of indices in empirical studies.

To examine how changes in the aggregation process affect regression results, we need indices that are identical in all aspects except their numerical form or their aggregation function. We address this issue by leaving the raw data unchanged when comparing regression results. In our baseline analysis, we use ten regime characteristics that are available for 186 countries and the period from 1919 up to 2018. Several robustness checks suggest that our results hold when we use other regime characteristics.

We show regression results on the relationship between democracy and economic development to illustrate the empirical consequences of changes in the design of an aggregation process. The main reason for this choice is that the related literature proposes different identification strategies. In this paper, we focus on the most common approaches and distinguish between OLS and 2SLS regressions. We make this distinction since these two regression techniques deal differently with classical measurement error (see Angrist and Pischke, 2009).[1]

[1]Classical measurement error produces an attenuation bias in an OLS estimate, whereas a 2SLS

We apply six standard aggregation methods to our regime characteristics.[2] To explore how these indices behave in empirical analyses, we run multiple regressions with each of them. Our results show that the choice of the aggregation technique significantly affects the magnitude of OLS estimates of the effect of democracy on economic development. We also show that the differences in the estimated effects persist if we address endogeneity problems with an instrumental variable approach. This result is remarkable because many scholars believe that indicators differ only regarding the degree to which they suffer from classical measurement error. Our findings imply that this presumption is unfounded because 2SLS estimates are not affected by classical measurement error. As an alternative, we offer an explanation suggesting that the choice of the aggregation method systematically influences the estimated levels of democracy of the regimes at the lower and upper end of the distribution. A simple econometric model and the results of some plausibility tests support our arguments.

To examine whether the decision regarding the scale of the index matters, we require indices that differ in their numerical form and are otherwise identical. We use two different methods to create such indices. The first method is an extended version of the machine learning approach developed by Gründler and Krieger (2016), while the other method transforms a continuous index by defining a threshold that splits regimes into two groups. For the machine learning approach, we find that a dichotomous index produces smaller OLS estimates than a continuous index. We show that this difference arises since dichotomous indices suffer more from classical measurement error and thus produce a greater attenuation bias. In line with this explanation, we find no statistically significant differences between continuous and dichotomous indices when running 2SLS regressions. For our second approach, we also observe that the magnitude of the OLS estimate decreases when we replace a continuous with a dichotomous index. The respective consequences for the 2SLS estimates depend on the choice of the threshold value that assigns the regimes to either the group of autocracies or democracies.

estimate remains unbiased.

[2] The list of aggregation techniques includes: an additive approach, an item-response approach, a multiplicative approach, a Machine Learning approach, and two approaches that combine additive and multiplicative indices.

We contribute to the literature that deals with the measurement of democracy. Several studies suggest that choice of the democracy index affects the results of regression analyses (see Casper and Tufis, 2003, Cheibub et al., 2010, Doucouliagos and Ulubaşoğlu, 2008, Gründler and Krieger, 2016, Krieckhaus, 2004), but only a few of them explore where the differences in the regression results come from. Knutsen and Wig (2015) suggest that conceptual differences play a role, while Elkins (2000) provides evidence that the scale of an indicator has a significant impact on an OLS estimate.[3] We are not aware of any study that investigates how regression results react to a change in the aggregation method.

Our paper also contributes to the literature that examines the relationship of democratic institutions and economic growth (see Acemoglu et al., 2019, Flachaire et al., 2014, Knutsen, 2015, Madsen et al., 2015, Murtin and Wacziarg, 2014, Persson and Tabellini, 2006, 2008, 2009, Rodrik and Wacziarg, 2005). We are the first who illustrate how methodological changes in the measure of democracy affect empirical results on the democracy-growth-nexus. Our findings imply that the choice of the aggregation technique does not affect the sign of the estimated effect of democracy on economic development, but significantly affects the size of the estimated growth effect.

We believe that our findings are relevant for many empiricists because indices belong to the standard tool kit of all social scientists. We are convinced that the choice of the numerical form and the choice of the aggregation function have an impact in various situations and thus urge scholars to be cautious when designing indices. Our paper helps to anticipate the consequences of decisions that must be made when designing the aggregation process.

This paper is organized as follows. Section 2 presents the basic ingredients of our democracy indices and provides an overview about existing data aggregation techniques. Section 3 describes our empirical framework. Section 4 shows how the choice of the aggregation tool affects regression results. Section 5 compares the performance of continuous and dichotomous indices. Section 6 concludes.

[3]Our analysis of the role of the numerical form of an index differs from Elkins' (2000) analysis in four ways. First, Elkins (2000) does not use the machine learning tool. Second, Elkins (2000) only reports results from OLS regressions (which are consistent with our results). Third, we show a greater number of robustness checks. Finally, we provided a detailed explanation for our results.

2 The measurement of democracy

The usual procedure for creating a democracy index includes three steps (Munck and Verkuilen, 2002): the first step defines the term "democracy" (*conceptualization*), the second step presents the regime characteristics that reflect the components of the definition (*operationalization*), and the third step specifies the rule that transforms the regime characteristics into an index (*aggregation*). Put differently, the level of democracy (Δ) is a function of the regime characteristics ($\mathbf{x}$):

$$\Delta^{\mathfrak{C}} := \mathfrak{F}^{\mathfrak{C}}(\mathbf{x}) = \mathfrak{F}^{\mathfrak{C}}(x_1, \ldots, x_m) \qquad \text{with} \qquad \mathbf{x} \in [0,1]^m \tag{1}$$

where $\mathfrak{F}^{\mathfrak{C}}\colon [0,1]^m \to \mathcal{D}$ denotes the aggregation function and $\mathfrak{C}$ the chosen concept of democracy.

2.1 Conceptualization

Goertz (2006) and Munck and Verkuilen (2002) argue that one must address two conceptual issues when designing a democracy index: (i) what are the *aspects* (or *dimensions*) that are associated with "democracy", and (ii) how do these aspects interact with each other. For both questions, many plausible answers exist (Teorell et al., 2019). We do not take sides for any of these answers, but show how the decisions made when creating an index influence the behavior of these indices in empirical analyses.

2.1.1 Aspects of democracy

Which institutional aspects should be included into a concept of democracy is the subject of a ongoing debate (Blaug and Schwarzmantel, 2016). Following O'Donnell (2001), we distinguish below between *narrow*, *realistic*, and *broad* concepts. Narrow concepts are focused on whether public elections for political mandates are competitive (see Przeworski, 1991, Schumpeter, 1942). Realistic concepts also require universal suffrage and basic political rights (see Dahl, 1971), whereas broader concepts also incorporate a wide range of other institutional factors (see Merkel, 2004).

Munck and Verkuilen (2002) argue that any theoretical debate on the choice of the institutional aspects is pointless because clear evaluation standards are lacking. We share this view. However, from an empirical point of view, narrow and broad concepts of democracy may create problems (see Bjørnskov and Rode, 2019, Munck and Verkuilen, 2002). Broad concepts are usually difficult to operationalize because of insufficient data availability and overlap with other economic concepts, such as corruption, economic freedom, and social inequality. Narrow concepts often do not sufficiently differentiate between different types of political regimes. In our main analysis, we therefore work with a realistic concept that consists of three aspects: *political participation, political competition*, and *freedom of opinion*.

2.1.2 Interaction between the aspects of democracy

The literature presents two theories on how individual aspects of democracy can interact with each other (Teorell et al., 2019). Some scholars argue that each aspect constitutes a necessary condition of democracy (Boix et al., 2013, Goertz, 2006). An argument that justifies this approach is that participation rights are meaningless if citizens cannot choose between candidates (or parties) with different policy programs. Similarly, freedom of expression might not play a role if no elections take place.

The alternative theory suggests that the chosen aspects of democracy are (partial) substitutes (Bollen, 1980, 1990, Treier and Jackman, 2008). The justification for this approach is that all aspects correlate with each other, and thus constitute a set of (partially) interchangeable "symptoms" of democracy (Teorell et al., 2019).

From a theoretical point of view, neither of the two theories is superior to the other since both approaches have some merits and objective evaluation criteria do not exist. Providers of indices should nonetheless pay some attention to this matter because the decisions that they make at this stage of the creation process affect the choice of the aggregation method (for details, see Section 2.3). Since the purpose of our paper is to examine how changes in the aggregation procedure influence the results of empirical analyses, we need to continue with both theories. Otherwise, we would not be able to compare the performance of all commonly used aggregation methods because most of them are simply not compatible with both theories.

2.2 Operationalization

We use ten regime characteristics that are available for a comprehensive sample of country-years to operationalize our aspects of democracy. To meet the guidelines proposed by Munck and Verkuilen (2002), we only use disaggregated data and draw our information from both objective and subjective sources.

We define *political participation* as citizens' right to elect their political rulers and representatives (Dahl, 1971). Suffrage might be limited, either through constitutional restrictions that exclude citizens because of their gender, race, or income, or by non-constitutional restrictions that result from material law, civil war, and repression. To capture the extent of constitutional restrictions, we use data from the Varieties of Democracy (V-Dem) database on the share of the adult people with legally granted suffrage (Coppedge et al., 2019). Measuring non-constitutional disenfranchisement is more difficult because of insufficient data availability. We address this issue by collecting data on voter turnout and calculating the voter-population ratio (see also Vanhanen, 2000).[4] These regime characteristics are equal to 0 if citizens have (de facto) no chance to elect their government. In addition, if a government suppresses opposition parties such that their supporters cannot participate in public elections, both the turnout rate and the voter-population ratio will be reduced. The main problem of using these two regime characteristics as measures for non-constitutional disenfranchisement is that low participation levels can also be caused by voluntary abstention. We still use these proxies since we believe that having only information on constitutional restrictions does not suffice to operationalize political participation.

Political competition exists if citizens with different party affiliations compete in public elections for political mandates (Przeworski, 1991). We operationalize this key aspect of democracy through five regime characteristics. The first characteristic is an expert-based index of party pluralism that discerns between five regime types.[5]

[4]We compile our data from a number of sources, including African Election Database (2019), Carr (2019), International Institute for Democracy and Electoral Assistance (2019), International Foundation for Electoral Systems (2019), Inter-Parliamentary Union (2019), Nohlen et al. (1999), Nohlen et al. (2001), Nohlen (2005), and Nohlen and Stöver (2010). A documentation of the collected data is available upon request.

[5]The five categories of the measure of party pluralism 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.

For this ordinal measure, we compile information provided by the V-Dem database (Coppedge et al., 2019), the databases of the Inter-Parliamentary Union (2019), and four election handbooks (Nohlen et al., 1999, Nohlen et al., 2001, Nohlen, 2005, Nohlen and Stöver, 2010). The four other regime characteristics are: (i) the share of votes not won by the strongest party/ candidate,[6] (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 (for a list of data sources, see Footnote 4).

The UN Human Rights Charter suggests that people enjoy *freedom of opinion* if they can freely decide on their sources of information and can express their political views even if these views are not compatible with the views of the government. To operationalize this aspect of democracy, we use gender-specific ratings on the freedom of debate from the V-Dem database (Coppedge et al., 2019).

2.3 Aggregation

Data aggregation consists of two parts: first, choosing the numerical form of the democracy indicator, and second, specifying the functional relationship between the regime characteristics and the level of democracy (Coppedge et al., 2011). Below, we provide an overview of the most common aggregation procedures and indicate for which concepts of democracy they are suitable.

2.3.1 The additive approach

Scholars who assume a concept of democracy with substitutable aspects often apply an additive aggregation function (Teorell et al., 2019):[7]

$$\Delta^{\text{add}} = \omega_1 \cdot x_1 + \ldots + \omega_m \cdot x_m \qquad \text{with} \qquad \sum_{j=1}^{m} \omega_j = 1 \tag{2}$$

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

[7] The list of additive measures of democracy includes (e.g.) the Polity index, the Freedom House indices, the Lexical Index of Electoral Democracy proposed by Skaaning et al. (2015), and the index of the Economist Intelligence Unit.

Table 1 Weighting scheme of additive and multiplicative index.

Regime Characteristic	Weight
Suffrage	0.0515
Voter-Population ratio	0.0599
Voter Turnout	0.0998
Party Pluralism	0.0878
Share of Votes	0.0690
Share of Parliamentary Seats	0.0643
Ratio Votes	0.0882
Ratio Parliamentary Seats	0.0827
Freedom of Discussion (female)	0.1961
Freedom of Discussion (male)	0.2009

Notes: This table reports the weights that we assign to the regime characteristics in the additive and multiplicative approach. To obtain these weights, we perform a Principle Component Analysis as suggested by Dreher (2006).

where $\omega_j \in (0,1)$ is the weight assigned to regime characteristic x_j. A reason for this choice is that additive indices implicitly assume that the regime characteristics affect the level of democracy independently from another and thus fit together with the conceptual assumptions.

The main challenge when implementing (2) is to specify the weights (ω_j). The classical approach is to assign the same weight to all regime characteristics (see e.g. the Polity index or the indices published by Freedom House). We argue that equal weighting is inappropriate for our data because our regime characteristics are not uniformly distributed among our three aspects of democracy (see also Munck and Verkuilen, 2002). Instead, we follow Coppedge et al. (2008), Dreher (2006), and Gygli et al. (2019) and use a Principle Component Analysis to determine the individual weights. We report these weights in Table 1.

Treier and Jackman (2008) argue that any indicator entails some degree of uncertainty and urge providers to inform the users about the extent of measurement uncertainty. Additive indices often fail to meet this requirement since (2) does not automatically produce confidence intervals. Another major concern against additive aggregation procedures is that the decision on how a particular regime characteristic influences the degree of democratization can hardly be grounded in theory and thus often appears to be arbitrary. For example, when using (2), one assumes that the marginal effect of each regime characteristic on the level of democracy is constant

(Treier and Jackman, 2008). The conceptual assumption of partial substitutability, however, does not imply this functional assumption because we would also achieve consistency between theory and aggregation procedure if the marginal effect of a regime characteristic on the degree of democratization varies with its own level.

By construction, (2) cannot be used to compute a dichotomous democracy index. To address this issue, scholars who prefer a dichotomous over a continuous index often define a threshold up to which a regime can be considered as a democracy. Bogaards (2012) and Cheibub et al. (2010) criticize this procedure for two reasons. First, it creates an inconsistency between theory and aggregation rule: while reaching a particular threshold constitutes a necessary condition, the conceptual assumptions associated with an additive index suggest that no necessary conditions exist. Second, any specific choice of a threshold value is arbitrary since it cannot be derived from theory.

2.3.2 The item-response approach

Item-response methods constitute an alternative procedure for scholars who prefer a concept of democracy with substitutable aspects (see Treier and Jackman, 2008, and Pemstein et al., 2010). The basic idea behind this approach is that democracy is a latent variable and can be modeled by the following data-generation process:

$$x_{rj} = \Delta_r + \varepsilon_{rj} \quad \text{with} \quad \varepsilon_{rj} \sim \mathcal{N}(0, \sigma_j^2) \tag{3}$$

where r denotes a regime, $j \in \{1, \ldots, m\}$ a regime characteristic, and Δ_r the true level of democracy. The parameters $\sigma_1^2, \ldots, \sigma_m^2$ indicate the error variances of the regime characteristics.

A practical challenge when applying an item-response approach is that all regime characteristics should have an ordinal scale with a finite number of categories (Treier and Jackman, 2008). Our regime characteristics do not meet this condition because seven of them have a continuous scale. To address this issue, we follow Pemstein et al. (2010) who define cutoffs to transform continuous into ordinal measures (for details, see Appendix Table C.1).[8] Below, $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_m)$ refers to the ordinal

[8]The estimation results reported in Section 4 do not significantly change if we use alternative cutoffs.

version of our regime characteristics, while $K = (K_1, \ldots, K_m)$ indicates the number of categories.

Another key assumption of item-response models is that the probability that a regime characteristic $\hat{x}_{rj}$ reaches a particular level $k \in \{1, \ldots, K_j\}$ can be expressed as follows (Pemstein et al., 2010):

$$Pr\left(\hat{x}_{rj} = k \,|\, \Delta_r, \alpha_j, \sigma_j\right) = \mathcal{F}\left(\frac{\alpha_{j,k} - \Delta_r}{\sigma_j}\right) - \mathcal{F}\left(\frac{\alpha_{j,k-1} - \Delta_r}{\sigma_j}\right) \tag{4}$$

where $\mathcal{F}(\cdot)$ denotes a cumulative distribution function and $\alpha_j = (\alpha_{j,1}, \ldots, \alpha_{j,K_j})$ a vector of unobserved thresholds for regime characteristic j. The likelihood for observing a particular data set is thus:

$$L\left(\mathrm{d}, \alpha, \sigma\right) = \prod_{r=1}^{N} \prod_{j=1}^{m} \left[\mathcal{F}\left(\frac{\alpha_{j,\hat{x}_{rj}} - \Delta_r}{\sigma_j}\right) - \mathcal{F}\left(\frac{\alpha_{j,\hat{x}_{rj}-1} - \Delta_r}{\sigma_j}\right)\right] \tag{5}$$

where N denotes the number of regimes in the sample, $\Delta = (\Delta_1, \ldots, \Delta_N)$, $\alpha = (\alpha_1, \ldots, \alpha_m)$, and $\sigma = (\sigma_1, \ldots, \sigma_N)$. Maximizing this likelihood with respect to all explanatory variables produces a democracy index for each regime.

Compared to the additive aggregation procedure, using an item-response approach has two main advantages (Treier and Jackman, 2008, Pemstein et al., 2010): first, the item-response model produces a distribution of indices for each regime and thus provides an opportunity to create measures of uncertainty,[9] and second, it does not require ad-hoc assumption about the marginal effects of the regime characteristics on the degree of democratization. Another feature of item-response approaches is that they produce a democracy index for a regime even if not all regime characteristics are available. However, Gründler and Krieger (2016) suggest that scholars should use this feature with caution since imbalanced regime characteristics can cause spurious changes in the predicted level of democracy. A similarity of the additive and the item-response approach is that both of them require the definition of a threshold value to produce dichotomous indices.

[9]The item-response approach usually creates confidence intervals that are smallest for regimes with an intermediate level of democracy (see Pemstein et al., 2010, Treier and Jackman, 2008). Teorell et al. (2019) suggest that this pattern is rather implausible since measurement uncertainty should be largest for hybrid regimes and smallest at the extremes.

2.3.3 Multiplicative approach

Providers of democracy indices usually apply a multiplicative aggregation method if they assume a concept in which a minimum of each aspect constitutes a necessary condition for democracy (Goertz, 2006):

$$\Delta^{\text{multi}} = x_1^{\omega_1} \cdot \ldots \cdot x_m^{\omega_m} \qquad \text{with} \qquad \omega_j \geq 0, \ \forall j \in \{1, \ldots, m\}. \tag{6}$$

This choice is consistent with the conceptual assumption because a multiplicative index exceeds 0 only if all regime characteristics are strictly greater than 0.

Similar to an additive index, the main difficulty when creating a multiplicative indicator is to assign the weights to the regime characteristics. For the sake of consistency, and since we do not find guidelines that advise against this approach, we adopt the weighing scheme designed for the additive index (see Table 1). Given our objective of showing the empirical consequences of using different aggregation methods, this choice appears to be appropriate, since if we set other weights, we cannot disentangle the role of the aggregation rule from the role of the weighting scheme in explaining regression results.

The multiplicative index also shares a methodological problem with the additive index because (6) does not create confidence intervals that indicate measurement uncertainty. A conceptual concern against (6) is that a multiplicative approach does not immediately follow from the assumption that a minimum of each aspect of democracy constitutes necessary condition. For example, taking the minimum of all regime characteristics is another aggregation method that is consistent with such a concept of democracy (Goertz, 2006).

A key difference between additive and multiplicative aggregation methods is that multiplicative techniques can directly produce dichotomous indices. However, (6) creates a dichotomous index only when all regime characteristics are dichotomous as well. Otherwise, one needs to specify threshold values required for a regime to be considered as democratic. Threshold values are, however, arbitrary (for a more detailed discussion, see Section 2.3.1).

2.3.4 Combining additive and multiplicative indices

Teorell et al. (2019) suggest that additive and multiplicative measures of democracy have their greatest discriminatory power at opposite ends of the spectrum: while additive indices vary greatly for autocratic regimes and relatively little for highly democratic regimes, multiplicative indicators differentiate more among democracies than among autocracies. To obtain an index with notable variation among both highly autocratic and highly democratic regimes, Teorell et al. (2019) calculate an average of an additive and a multiplicative measure of democracy:

$$\Delta^{\text{av}} = \lambda \cdot \Delta^{\text{add}} + (1 - \lambda) \cdot \Delta^{\text{multi}} \tag{7}$$

where $\lambda \in (0, 1)$ is the weight assigned to the additive index. We set $\lambda = 0.44$ according to the results of a Principle Component Analysis.[10]

Any concept assuming that particular aspects of democracy constitute necessary conditions is inconsistent with (7). The reason is that the degree of democratization can exceed 0 even if the necessary regime characteristics are equal to 0. As an alternative, scholars who prefer definitions with necessary conditions can apply a Cobb-Douglas function to combine additive and multiplicative indices:

$$\Delta^{\text{cobb}} = \left(\Delta^{\text{add}}\right)^{\lambda} \cdot \left(\Delta^{\text{multi}}\right)^{1-\lambda} \quad \text{with} \quad \lambda \in (0, 1). \tag{8}$$

The aggregation procedures described by (7) and (8) share three weaknesses with the additive and multiplicative approach. First, none of them automatically creates confidence intervals that indicate the degree of measurement uncertainty.[11] Second, the conceptual assumptions do not completely explain the shapes of the aggregation functions. Finally, the creation of dichotomous measures of democracy requires an arbitrary definition of threshold values.

[10] The estimation results that we present in Section 4 do not significantly change if we assign the same weight to the additive and multiplicative index (see Teorell et al., 2019).

[11] Teorell et al. (2019) address this issue with a rather complex approach that uses variation in the weighting schemes. In contrast to the confidence intervals produced by the item-response approach (see Treier and Jackman, 2008 and Pemstein et al., 2010), the approach proposed by Teorell et al. (2019) creates confidence intervals that are largest for hybrid regimes. Teorell et al. (2019) argue that their measures of uncertainty are more plausible.

2.3.5 Machine Learning approach

In an earlier study (see Gründler and Krieger, 2016), we proposed an aggregation procedure that is based on a Machine Learning technique for pattern recognition, known as Support Vector Machines (SVM). Our motivation for developing a new aggregation method was that the conventional approaches need specific assumptions about the shape of the aggregation function and that these assumptions are subject to severe criticism because of arbitrariness and simplicity (see Cheibub et al., 2010, Munck and Verkuilen, 2002). When applying SVM, we relax these assumptions and solve non-linear optimization problems to address the question of how to transform the regime characteristics into a democracy index. The downside of our method is that the shape of the aggregation function is not grounded in theory and lacks an explicitly representation.[12]

Since SVM is a supervised machine learning technique, its application requires a set of observations (henceforth: *priming data*) for which we observe both the input characteristics and the outcome variable (see Abe, 2005, Steinwart and Christmann, 2008). We proceed in two steps to satisfy this prerequisite. First, we argue that the level of democracy of the most and least democratic regimes is uncontroversial and that these regimes can thus be used as priming data. The motivation for this argument has been encapsulated by Lindberg et al. (2014) who wrote that "*almost everyone agrees that Switzerland is democratic and North Korea is not*" (for similar statements, see Cheibub et al., 2010 and Diamond, 2002). The second step uses the indices of Pemstein et al. (2010) and Teorell et al. (2019) to identify regimes whose degree of democratization is uncontroversial. In our basic specification, we label a country-year as a highly autocratic (democratic) regime if it belongs to the lower (upper) decile of either of the two indices.[13]

The aggregation process of our machine learning approach consists of four steps. In the first step, we randomly select some country-years from the priming data to

[12] Since we observe that the degree of democratization can exceed 0 if a regime characteristic is equal to 0, the machine learning procedure is not consistent with a theory in which the aspects of democracy constitute a set of necessary conditions.

[13] We test whether our machine learning indices significantly change when using other criteria. Since we only find minor changes, none of the regression results reported in Sections 4 and 5 depend on the choice of the labeling criteria.

produce a training set $\mathcal{T}_\eta$. In the second step, we use a classification/regression method from the SVM toolboox (for details, see Appendix B) and the training set $\mathcal{T}_\eta$ to estimate the aggregation function $\widehat{\mathfrak{F}}_\eta\colon [0,1]^m \to [0,1]$. The third step uses the estimated aggregation function $\widehat{\mathfrak{F}}_\eta(\cdot)$ to compute a democracy indicator for each country-year observation in our data set:

$$\Delta_{i,t,\eta} = \widehat{\mathfrak{F}}_\eta\left(x_{i,t,1}, \ldots, x_{i,t,m}\right)$$

where i indicates the country and t the year. In the last step, we repeat steps $1-3$ for all iterations $\eta \in \{0, \ldots, \eta_{max}\}$. Our aggregation method thus creates a distribution of indices for each country-year observation. We use the median of each distribution as democracy index and other percentiles as bounds of the confidence intervals.

The SVM toolbox includes classification and regression techniques (Steinwart and Christmann, 2008). We use these tools to compute (otherwise identical) continuous and dichotomous measures of democracy. In comparison to the other aggregation approaches, the machine learning procedure thus has the advantage of producing dichotomous indicators for all kinds of regime characteristics without requiring a manual definition of a threshold value. Another advantage of this approach is that it produces measures of uncertainty for both the continuous and the dichotomous indices.

The performance of the machine learning approach depends on the priming data because this data constitutes the basis on which the SVM techniques "learn" the functional relationship between the regime characteristics and the level of democracy. Theoretically, the priming data must meet two prerequisites: first, the country-years that are part of the priming data must be correctly labeled, and second, these observations need to reflect the institutional heterogeneity among the autocratic and democratic regimes.[14] Appendix A shows the results of different tests that indicate that our priming data satisfies both conditions.

[14]If the second condition is not satisfied, the machine learning approach may create indices that discriminate in favor of specific types of government. For example, if the priming data only includes autocracies in which no elections take place, the method may have difficulties to correctly classify autocratic regimes with non-competitive elections.

3 Empirical framework

We illustrate the empirical consequences of using different aggregation tools with regression results on the effect of democracy on economic growth. We choose this topic for two key reasons: First, previous studies report ambiguous results on the question of whether democracy causes long-run economic growth (Acemoglu et al., 2019, Doucouliagos and Ulubaşoğlu, 2008, Gründler and Krieger, 2016, Knutsen, 2015, Madsen et al., 2015, Murtin and Wacziarg, 2014, Papaioannou and Siourounis, 2008, Persson and Tabellini, 2006, Tavares and Wacziarg, 2001). Since the reasons for the heterogeneity of the regression results are still not clear, we believe that improving the understanding of the role of the democracy index is a useful contribution to this strand of literature. Second, the literature on the effect of democracy on economic development proposes different identification strategies. We exploit this variety to investigate how the consequences of using different aggregation methods depend on the applied regression technique.

Three endogeneity issues complicate an analysis of the effect of democracy on economic development (Acemoglu et al., 2019): first, autocratic regimes differ from democratic regimes in non-observable factors that also affect economic development, second, causality may run from economic development to democracy,[15] and third, democratization is often preceded by a temporal decline in GDP per capita. To address these endogeneity problems, many scholars estimate a dynamic fixed effect model:

$$Y_{i,t} = \sum_{l=1}^{L} \beta_l \cdot Y_{i,t-l} + \gamma \cdot D_{i,t} + \xi_i + \eta_t + \varepsilon_{i,t} \tag{9}$$

where D denotes the degree of democratization of country i in year t, Y the log of GDP per capita, ξ the country fixed effect, η the year fixed effect, and ε the error term.[16]

The dynamic fixed effect model correctly identifies the effect of democracy on economic development if the error term is uncorrelated with the level of democracy.

[15] For studies examining how economic development and economic shocks affect democratization, see Acemoglu et al. (2008), Aidt and Franck (2015), Aidt and Leon (2016), Brückner and Ciccone (2011), Cervellati et al. (2014), Gundlach and Paldam (2009), Lipset (1959), Murtin and Wacziarg (2014), and Przeworski (2000).

[16] Our data on GDP per capita comes from the Maddison Project Database 2018 (see Bolt et al., 2018).

Since this condition is unlikely to be satisfied due to omitted time-varying factors, some recent studies use a two-stage least squares (2SLS) approach in which the average degree of democratization in neighboring countries serves as the instrument for the domestic degree of democratization (see Acemoglu et al., 2019, Dorsch and Maarek, 2019, Persson and Tabellini, 2009):

$$D_{i,t} = \sum_{l=1}^{L} \delta_l \cdot Y_{i,t-l} + \alpha \cdot Z_{i,t} + \zeta_i + \tau_t + \iota_{i,t} \tag{10}$$

with

$$Z_{i,t} = \frac{1}{|\mathcal{R}|} \sum_{j \in \mathcal{R}} D_{j,t} \quad \text{and} \quad \mathcal{R} = \{j : j \neq i, r_j = r_i\} \tag{11}$$

where r_i denotes the region in which country i is located.[17] The motivation for this instrumentation strategy is that transitions from autocracy to democracy (and vice versa) often occur in regional waves (Huntington, 1993, Teorell, 2010).

Our 2SLS approach produces unbiased estimates for the effect of democracy on economic development if two assumptions hold (Angrist and Pischke, 2009): first, the regional degree of democratization (Z) and the domestic degree of democratization (D) correlate with each other, and second, the regional level of democracy affects economic development only through its effect on the domestic level of democracy. The latter assumption may be violated for several reasons. Changes in the regional degree of democratization may, for example, have an effect on the regional level of political stability, and thus influence domestic prices, investments, and trade flows.

Acemoglu et al. (2019) and Dorsch and Maarek (2019) allay concerns against the validity of the exclusion restriction with several tests. Since our purpose is not to identify the causal effects of democratization events, we only repeat some of these tests and are cautious with causality claims regarding the relationship of democracy and economic development. For our research question, a violation of the exclusion restriction is unproblematic since the resulting bias in the coefficient estimates is the same for all democracy indices and thus does not depend on the aggregation technique.

[17] In our baseline analysis, we use the classification of the United Nations to divide the world into 19 regions. Results for other classification schemes look similar and are available upon request.

4 Differences in the shape of the aggregation function

We use several continuous measures of democracy to illustrate how the functional assumptions about the relationship of the regime characteristics and the degree of democratization affect the results of OLS and 2SLS regressions. We proceed in two steps: first, we compare the empirical performance of continuous indices that only differ regarding their aggregation functions and second, we provide a detailed explanation for why the design of the aggregation method significantly influences regression results.

4.1 Estimation results

4.1.1 OLS estimates

In Table 2, we show OLS results from estimating Equation (9) with an unbalanced panel that covers 163 countries and covers the period from 1919 to 2016. The only difference between the six columns is that we used different aggregation tools to produce the democracy index. In all regressions, we add four lags of the dependent variable to our model and cluster the standard errors at the country level.

Column 1 uses the machine learning index. In line with other recent studies that exploit fixed effect models, we find a positive and statistically significant relationship between democracy and economic development. Our OLS estimate suggests that a transition from autocracy ($D = 0$) towards democracy ($D = 1$) increases GDP per capita by 1.7 percent per year.[18] The estimated long-run effect is 113 percent and thus lies between the cumulative long-run effects reported by Acemoglu et al. (2019) and Madsen et al. (2015).[19]

[18]Compared to the OLS results reported by Papaioannou and Siourounis (2008) and Acemoglu et al. (2019), we find a larger estimate. In Section 5, we will show that this difference can be largely explained by differences in the numerical form of the democracy index. The estimates reported by Madsen et al. (2015) who use the Polity index (i.e. a quasi-continuous index with an additive aggregation function) are larger than the estimate reported in Column 1 of Table 2.

[19]Below, we only compare the performance of different aggregation methods with respect to the short-run effect (captured by the parameter γ). A concern against this focus may be that it is theoretically unclear whether using different aggregation techniques has the same consequences for the short-run and the long-run effect. If the choice of the aggregation method affects the estimates of the lagged dependent variable ($\beta_1, \ldots, \beta_4$), we would find different consequences for the estimates of the short-run and the long-run effect of a democratic transition. To alleviate this concern, we report the estimates of the parameters $\beta_1, \ldots, \beta_4$ in our baseline table. We observe that these estimates are virtually the same in all six columns. The difference in the estimated long-run effects can thus be fully explained with the difference in the estimates of the parameter γ.

Table 2 Consequences of using different aggregation functions — OLS estimates

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.017***	0.032***	0.033***	0.023***	0.027***	0.022***
	(0.003)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Income_{t-1}	1.185***	1.184***	1.184***	1.184***	1.184***	1.184***
	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
Income_{t-2}	-0.108	-0.108	-0.108	-0.108	-0.108	-0.108
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Income_{t-3}	-0.084**	-0.085**	-0.085**	-0.084**	-0.084**	-0.084**
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
Income_{t-4}	-0.008	-0.008	-0.007	-0.008	-0.008	-0.007
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Observations	10,026	10,026	10,026	10,026	10,026	10,026
Countries	163	163	163	163	163	163
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985
Equal. (p-val.)	–	0.000	0.001	0.087	0.003	0.144
Long-run effect	1.133	1.972	2.057	1.446	1.674	1.406

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

In Column 2, we use the additive index. We observe that the change in the aggregation method increases the OLS estimate from 0.017 to 0.032. The result of the Wald test indicates that the difference between the regression coefficients is statistically significant at the 1 percent level. We find a similar result when we use the item-response approach instead of the additive approach (see Column 3).

Column 4 presents the result for the multiplicative indicator. Compared to the machine learning index, the OLS estimate of the effect of democracy on economic development increases significantly, but to a lesser extent than the additive index. The difference between the regression coefficients produced by the additive measure and the multiplicative measure is statistically significant at the 5 percent level (p-value: 0.012).

Column 5 uses the index that is a weighted average of the additive and the multiplicative indicator. We observe that the OLS estimate produced by this index is between the estimates of the underlying indices and statistically different from the estimate produced by the machine learning index. Column 6 shows that the estimation result changes when we apply a Cobb-Douglas function to combine the additive and the multiplicative measure. In this case, we obtain a slightly smaller

OLS estimate than with the multiplicative index.

In sum, Table 2 suggests that the choice of the aggregation method affects the results of OLS regressions in a significant manner. We find that the size of the estimated effect of democracy on economic development increases by more than 90 percent when we switch from the aggregation method that produces the smallest regression coefficient (see Column 1) to the aggregation technique that produce the largest regression coefficient (see Column 3). In the next section, we investigate whether the differences in the estimates disappear when we use an instrumental variable approach.

4.1.2 2SLS estimates

To investigate the consequences of applying different aggregation methods for the results of 2SLS regressions, we would ideally like to use an instrumental variable that does not depend on the design of the aggregation method. Unfortunately, the literature does not provide such an instrument. We thus exploit the regional (jack-knifed) degree of democratization (see Acemoglu et al., 2019, Persson and Tabellini, 2009). To ensure that the instrument does not change when we switch from one aggregation method to another, we compute the mean of the instruments produced by our six indices:[20]

$$Z_{i,t} = \frac{1}{6} \cdot \left(Z_{i,t}^{ML} + Z_{i,t}^{Add} + Z_{i,t}^{IR} + Z_{i,t}^{Multi} + Z_{i,t}^{AM_{Av}} + Z_{i,t}^{AM_{CD}} \right). \tag{12}$$

Table 3 presents the results of our 2SLS regressions. In Column 1, we use the machine learning index. Compared to the corresponding OLS estimate reported in Column 1 of Table 2, we observe an increase in the regression coefficient (see Panel A). This increase is consistent with other studies that use the regional degree of democratization as an instrumental variable (see Acemoglu et al., 2019). Also in line with previous studies, we find a strong first-stage relationship between the regional degree of democratization and the domestic degree of democratization.

Column 2 replaces the machine learning index with the additive index. Two

[20]We also run regressions in which a change in the aggregation method causes a change in the instrumental variable. Appendix Table C.2 presents these results. We observe that second-stage estimates hardly change compared to the estimates reported in Table 3.

Table 3 Consequences of using different aggregation functions — 2SLS estimates

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second-stage estimates					
Democracy	0.032***	0.055***	0.067***	0.039***	0.045***	0.038***
	(0.006)	(0.010)	(0.012)	(0.007)	(0.008)	(0.007)
Income$_{t-1}$	1.181***	1.180***	1.177***	1.180***	1.180***	1.180***
	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
Income$_{t-2}$	-0.106	-0.106	-0.105	-0.106	-0.106	-0.106
	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)	(0.070)
Income$_{t-3}$	-0.084**	-0.084**	-0.084**	-0.084**	-0.084**	-0.084**
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
Income$_{t-4}$	-0.009	-0.009	-0.008	-0.009	-0.009	-0.009
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Equal. (p-val.)	–	0.000	0.000	0.273	0.038	0.337
	Panel B: First-stage estimates					
Demo. (reg.)	0.943***	0.554***	0.453***	0.785***	0.684***	0.802***
	(0.086)	(0.051)	(0.042)	(0.072)	(0.062)	(0.073)
Equal. (p-val.)	–	0.000	0.000	0.069	0.003	0.105
Observations	10026	10026	10026	10026	10026	10026
Countries	163	163	163	163	163	163
SW (F-stat.)	118.91	117.72	116.05	119.52	121.10	119.38
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.780	2.828	3.322	2.121	2.381	2.089

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\, p < 0.10$, $^{**}\, p < 0.05$, $^{***}\, p < 0.01$.

consequences are striking: first, the second-stage estimate increases from 3.2 to 5.5 percent, and second, the first-stage estimate decreases from 0.943 to 0.554. Both changes are statistically significant at the 1 percent level. Column 3 shows that the first-stage/second-stage estimate further decreases/increases when we use the item-response approach to create a continuous measure of democracy.

Column 4 reports the regression results for the multiplicative index. We observe that this indicator produces first- and second-stage estimates that lie between the estimates of the machine learning indicator and the additive indicator. A notable difference compared to the OLS estimates reported in Table 2 is that the Wald test does not indicate a significant difference in the estimated effect of democracy on economic development between the machine learning index and the multiplicative index. The reason for this change is that the standard errors of the second-stage estimates are larger than the standard errors of the OLS estimates.

In Column 5, we use the index that is a weighted mean of the additive and the multiplicative indicator. This index produces a second-stage estimate of 0.045 and a first-stage estimate of 0.684. These estimates differ significantly from the estimates reported in Columns 1 and 3, but not much from the estimates reported in other columns. Column 6 uses a Cobb-Douglas function to combine the additive and the multiplicative index. Consistent with our OLS results, we observe that this index produces similar estimates as the multiplicative index.

Taken together, Table 3 shows that a change in the aggregation method can create significant changes in the results of 2SLS regressions. This result is notable because many scholars examining the consequences of political transitions suggest that democracy indicators differ mainly with regard to their amount of classical measurement error and that these indices thus produce similar first- and second-stage estimates. In Section 4.2, we will explain why this logic does not necessarily apply and develop an alternative explanation. However, before we turn to the next section, we present the results of some robustness checks to allay the concern that our baseline results have weak external validity.

4.1.3 Robustness checks

Many economists argue that annual data is inappropriate for studying the causes of long-run economic growth. These scholars rather prefer data that is averaged over multiple years because data averaging filters out business cycle fluctuations and mitigates the role of measurement error in the explanatory variables (Durlauf et al., 2005). Appendix Tables C.3 and C.4 suggest that the choice of the aggregation method also affects the results of OLS and 2SLS regressions when we use 5-year averaged data rather than annual data. However, the differences in the regression coefficients are slightly less pronounced.

In Appendix Table C.5 and C.6, we extend our regression models by control variables that are available for our sample period (population growth, civil conflict, rule of law).[21] The results show that the inclusion of these covariates has only a

[21] We use data from Brecke (1999) and the Uppsala Conflict Data Program to create a binary index of civil conflict. The measure of the rule of law comes form the V-Dem database. The data on population growth is obtained from four sources: Bolt et al. (2018), the Cross-National Time Series Data Archive, the World Bank, and the web page `www.populstat.info`.

small impact on the consequences of using different aggregation methods.

Appendix Tables C.7 – C.10 show results from regressions in which the average years of schooling and an expert-based measure of private property rights serve as the outcome variables.[22] In line with other empirical studies, we find that both the education level and the quality of the economic institutions increases in the degree of democratization (see De Haan and Sturm, 2003, Harding and Stasavage, 2013). We also observe that the differences between our measures of democracy persist.

Political scientists controversially debate about which institutional aspects should be part of a concept of democracy. In our baseline analysis, we chose a concept that includes three aspects (political participation, political competition, freedom of opinion) and resembles the concept of Dahl (1971). To illustrate that the impact of the aggregation technique on the results of OLS and 2SLS regressions does not significantly change if we change the institutional aspects, we use two alternative concepts. In the first alternative concept, we only consider the aspect of political competition (see Przeworski, 1991), while we use *judiciary independence* as a fourth institutional aspect in the second alternative concept of democracy (see O'Donnell, 2001). Appendix Tables C.11 – C.14 show that our findings are robust to these conceptual changes.

Another concern may be that we strategically selected our regime characteristics and that the differences between the aggregation methods disappear when we use alternative regime characteristics. To allay this legitimate concern, we repeat our baseline analysis with the regime characteristics proposed by Teorell et al. (2019). Appendix Tables C.15 and C.16 illustrate that this change in the set of regime characteristics does not changes our estimation results.

In sum, the findings of our robustness checks confirm that the choice of the aggregation function significantly affects the size of OLS and 2SLS estimates. Our results also suggest that the rank order of the estimates is fairly robust. In all regressions, we find that the machine learning index indicates the smallest effects. The largest estimate is always either produced by the additive index or the index that is based on the item-response approach. For the indices that we obtain from

[22] The education data comes from Barro and Lee (2013). The V-Dem database serves as source for the information about private property protection.

combining the additive and the multiplicative index, we observe that using a Cobb-Douglas function leads to smaller estimates than taking a weighted average.

4.2 Explanation

4.2.1 Theory

When democracy indices produce OLS estimates that differ in their magnitude, the usual explanation is that these indices suffer differently from classical measurement error and thus create different attenuation biases. We doubt that this explanation fits to our estimation results because it contradicts with the results from our 2SLS regressions: if the different attenuation biases explained the differences in the OLS estimates, we would not find any statistically significant difference across the 2SLS estimates since an instrumental variable approach fully addresses the problem of classical measurement error (see Angrist and Pischke, 2009). As an alternative, we propose an explanation that assumes systematic differences between indicators for autocratic and democratic observations (we substantiate this assumption in Section 4.2.2). The following simple model shows that this assumption suffices to explain both the differences in the OLS estimates and the differences in the 2SLS estimates.

Assume that the degree of democratization (D) influences an observable outcome variable (Y) in the following manner:

$$Y_i = \alpha + \beta \cdot D_i + \varepsilon_i \tag{13}$$

where $\alpha, \beta > 0$ indicate unknown parameters and ε a randomly distributed error term. For analytical convenience, we also assume that m of the n independent observations have a level of democracy of D_{low} and that the level of democracy of the remaining $n - m$ observations is $D_{high} > D_{low}$.

Consider now two democracy indicators and suppose that the first measure (Δ^1) indicates a lower (higher) degree of democratization than the second measure (Δ^2) for regimes with a low (high) level of democracy:

$$\Delta_j^2 = \Delta_j^1 + \mathcal{E}(D_j) \quad \text{for} \quad j \in \{low, high\} \tag{14}$$

with

$$\mathcal{E}\big(D_j\big) \;=\; \begin{cases} -\,\eta & \text{for} \quad D_j \;=\; D_{high} \\ \gamma & \text{for} \quad D_j \;=\; D_{low} \end{cases} \qquad \text{with} \quad \eta \,>\, 0 \quad \text{and} \quad \gamma \,>\, 0.^{23} \tag{15}$$

When using these indicators as proxies for the (unknown) level of democracy (D), we obtain the following OLS estimates of the effect of democracy on the outcome variable (Y):

$$\widehat{\beta}^{k}_{ols} \;=\; \frac{\operatorname{cov}\big(Y,\,\Delta^k\big)}{\operatorname{var}\big(\Delta^k\big)} \;=\; \frac{m\cdot\sum_{i=m+1}^{n}(Y_i-\bar{Y})-(n-m)\cdot\sum_{i=1}^{m}(Y_i-\bar{Y})}{(\Delta^k_{high}-\Delta^k_{low})\cdot m\cdot(n-m)} \tag{16}$$

where $k \,\in\, \{1,\,2\}$ indicates whether we apply the first or the second democracy indicator. Equation (16) shows that the magnitude of the OLS estimate increases when the difference between Δ^k_{high} and Δ^k_{low} decreases. The second indicator thus produces larger OLS estimates than the first index:

$$\Delta^2_{high} \,-\, \Delta^2_{low} \,<\, \Delta^1_{high} \,-\, \Delta^1_{low} \quad\Rightarrow\quad \widehat{\beta}^{2}_{ols} \,>\, \widehat{\beta}^{1}_{ols}. \tag{17}$$

To show that the second index also creates a larger second-stage estimate and a smaller first-stage estimate in a 2SLS regression, we assume that we have an observable variable $Z \,\geq\, 0$ which positively correlates with the level of democracy (D) and influences the outcome variable (Y) only through its effect on the level of democracy. When we use Z as an instrumental variable in a 2SLS regression, we obtain the second-stage estimates:

$$\widehat{\beta}^{k}_{iv} \;=\; \frac{\operatorname{cov}\big(Y,\,Z\big)}{\operatorname{cov}\big(\Delta^k,\,Z\big)} \;=\; \frac{\widehat{\delta}_{ols}}{\widehat{\rho}^{k}_{ols}}, \tag{18}$$

where $\widehat{\rho}^{k}_{ols}$ is the OLS estimator of the first-stage model:

$$\Delta^k \;=\; \pi \,+\, \rho\,\cdot\, Z_i \,+\, \xi_i \quad \text{with} \quad \rho \,>\, 0, \tag{19}$$

and $\widehat{\delta}_{ols}$ denotes the OLS estimator of the reduced-form model:

[23]Below, we only consider cases in which $d_{2,high} \,=\, \mathrm{d}_{high} \,-\, \eta \,>\, \mathrm{d}_{low} \,+\, \gamma \,=\, d_{2,low}$.

$$\mathrm{Y}_i = \zeta + \delta \cdot \mathrm{Z}_i + \iota_i \quad \text{with} \quad \delta > 0. \tag{20}$$

From

$$\begin{aligned}
\widehat{\rho}^{k}_{ols} &= \frac{\operatorname{cov}\left(Z, \Delta^k\right)}{\operatorname{var}\left(Z\right)} \\
&= \frac{\frac{1}{n} \cdot \left(\Delta^k_{high} - \Delta^k_{low}\right) \cdot \left(m \cdot \sum_{i=m+1}^{n}(Z_i - \bar{Z}) - (n-m) \cdot \sum_{i=1}^{m}(Z_i - \bar{Z})\right)}{\operatorname{var}\left(\mathrm{Z}\right)}
\end{aligned}$$

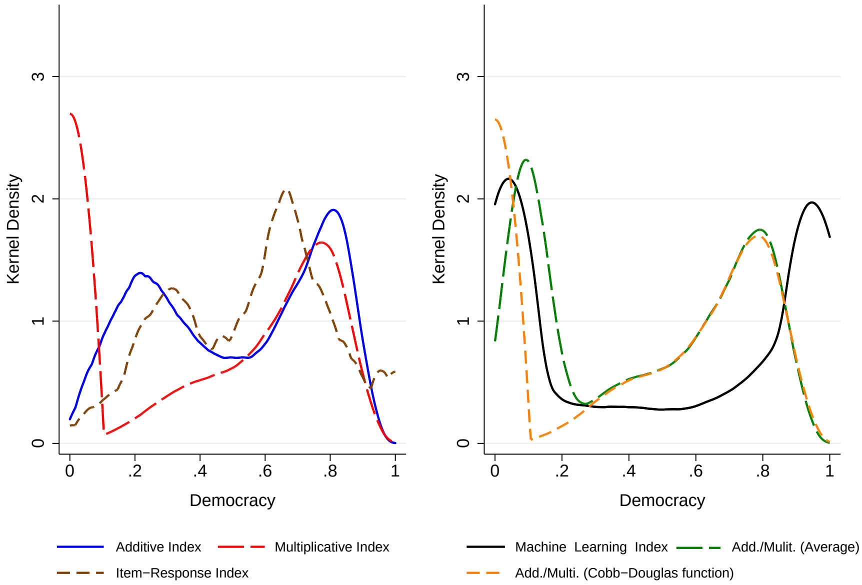
we can infer that the first-stage estimate increases if we replace the first index with the second index. The reason is once again that the second measure indicates a smaller difference between the regimes with a high and low level of democracy. A direct consequence of the difference in the first-stage estimates is that the second indicator produces a larger second-stage estimate than the first indicator:

$$\Delta^2_{high} - \Delta^2_{low} < \Delta^1_{high} - \Delta^1_{low} \quad \Rightarrow \quad \widehat{\rho}^{2}_{ols} < \widehat{\rho}^{1}_{ols} \quad \Rightarrow \quad \widehat{\beta}^{2}_{iv} > \widehat{\beta}^{2}_{iv}. \tag{21}$$

4.2.2 Supporting evidence

Our simple econometric model assumes that our democracy indices systematically differ from each other with respect to the level of democracy that they indicate for regimes at the lower and upper end of the distribution. To substantiate this assumption, we present the results of kernel density estimations (see Figure 1). We observe that all density functions are bimodal and have local maxima in the lower and upper part of the spectrum. However, the exact locations of the maxima differ considerably from each other. For example, the density function of the machine learning index has a lower maximum at $\Delta \approx 0.05$ and a upper maximum at $\Delta \approx 0.95$, while the density function of the additive index has its maxima at $\Delta \approx 0.2$ and $\Delta \approx 0.80$. These differences in the location of the maxima suggest that the differences between autocratic and democratic regimes are less pronounced when we consider the additive rather than the machine learning indicator. According to our theory, the additive indicator should thus produce larger OLS and 2SLS estimates than the machine learning index. The estimation results presented in Section 4.1 show that this prediction holds.

Figure 1 Kernel densities



Notes: The figure shows the kernel densities of our six democracy indices. We use the Epanechnikov kernel to estimate the density functions.

Another way to support the key assumption of our model is to examine how our six continuous indicators react when a political transition from autocracy towards democracy (or vice versa) occurs. If our assumption holds, we should find two patterns: first, the indicated change in the level of democracy varies considerably across our six indices, and second, greater changes coincide with smaller OLS and 2SLS estimates on the effect of democracy on economic development. To check whether these patterns exist, we exploit the dichotomous indices of Bjørnskov and Rode (2019), Boix et al. (2013), and Papaioannou and Siourounis (2008) and create measures that reflect the average change in the degree of democratization during a transition:

$$\Theta_k^j = \frac{1}{|\mathcal{S}_k|} \sum_{(i,t)\,\in\,\mathcal{S}_k} |\,\Delta_{i,t}^j - \Delta_{i,t-1}^j\,| \tag{22}$$

where $\mathcal{S}_k$ denotes the set of all regime changes either indicated by Bjørnskov and Rode (2019), Boix et al. (2013), or Papaioannou and Siourounis (2008) and $\Delta \in [0, 1]$ the degree of democratization indicated by the continuous indicator j. We present these measures (Θ) in Table 4. As expected, we find notable differences between our continuous indicators. We also see that these differences and the differences in

Table 4 Average changes in the level of democracy during political transitions

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
Change (BR)	0.289	0.168	0.184	0.243	0.210	0.249
Change (BMR)	0.386	0.219	0.239	0.328	0.280	0.336
Change (PS)	0.363	0.196	0.193	0.278	0.241	0.276

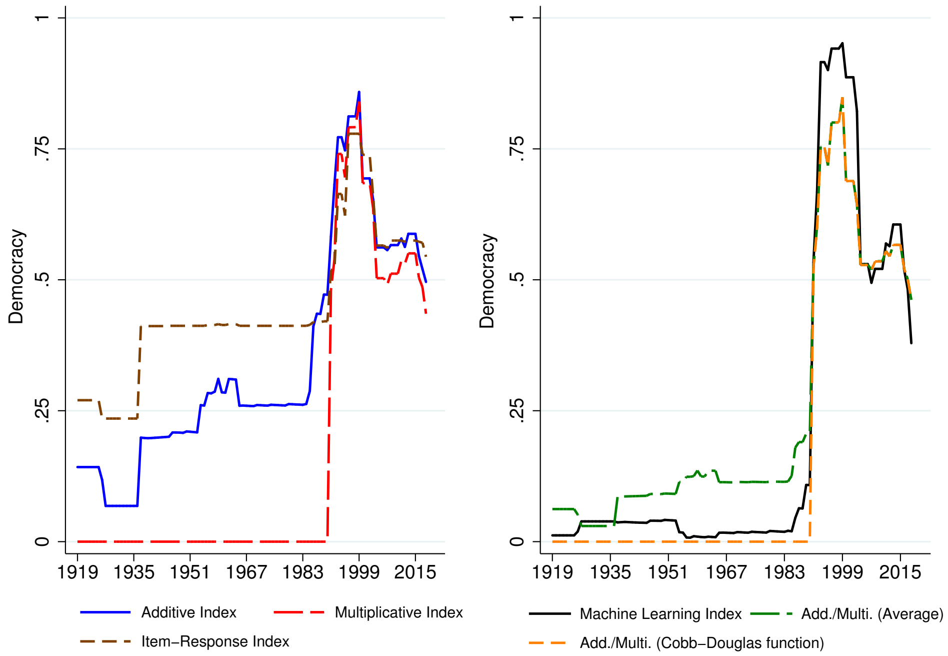
Notes: This table reports results from estimating Equation (22). The figures show how much a democracy index change, on average, when Bjørnskov and Rode (2019), Boix et al. (2013), or Papaioannou and Siourounis (2008) indicate a political transition from autocracy to democracy (or vice versa).

the regression results relate to each other in the predicted manner. For example, the machine learning indicator shows the greatest response to a political transition and suggests the smallest effect of democracy on economic development, while the additive and the item-response index change relatively little and produce relatively large estimates. Our model also fits if we compare the indices that we create by combining the additive and multiplicative index because using a weighted average leads to smaller changes in the level of democracy and larger regression coefficients than using a Cobb-Douglas function.

Since our indicators differ only with respect to their aggregation techniques, an obvious question is whether we can explain the patterns observed in Figure 1 and Table 4 with differences in the assumed functional relationship between the regime characteristics and the degree of democratization. To show that this is indeed the case, we consider the indices of two countries in greater detail.

In Figure 2, we show the level of democracy in the Soviet Union and the Russian Federation for different aggregation methods. We observe that all indices indicate a distinct increase in the degree of democratization after the collapse of the Soviet Union in 1991. We think this increase is plausible because fairly free multi-party electios took place in the early years of the Russian Federation, while single-party elections were held in the Soviet period (Nohlen and Stöver, 2010, Sakwa, 2005). We also see that all indices decrease after the inauguration of Vladimir Putin. Several expert reports confirm the plausibility of this decrease (see Hale et al., 2004, Sakwa, 2010). Major differences between the six indices mainly exist for the Soviet period: while the machine learning index, the multiplicative index, and the index that we obtain by combing the additive and the multiplicative index with a Cobb-Douglas

Figure 2 Democracy in the Soviet Union and the Russian Federation (1919 – 2018).



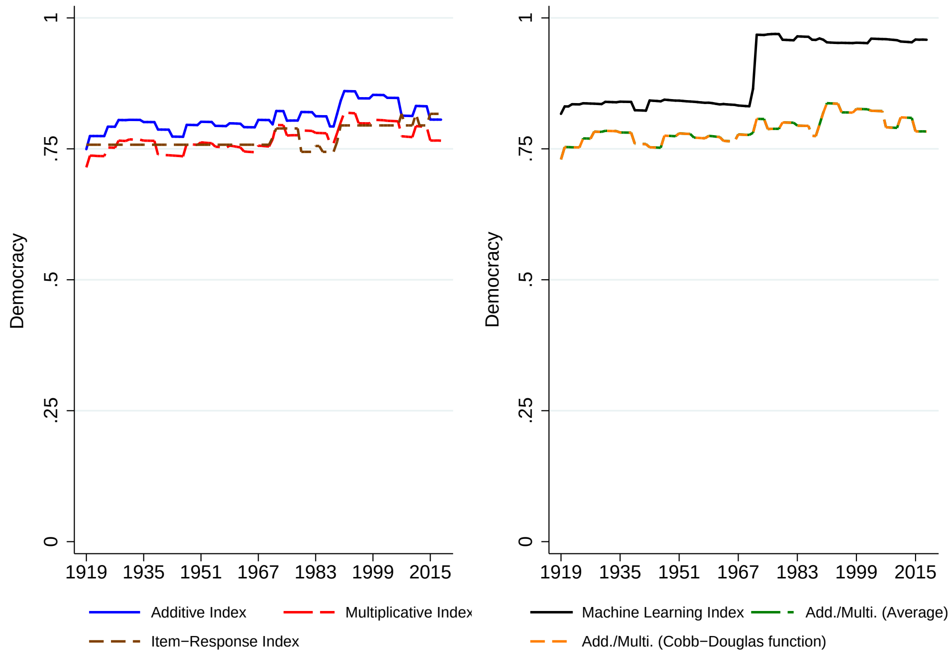
Notes: The figures show the level of democracy of the Russian Federation and the Soviet Union, depending on how we aggregate our ten regime characteristics. From 1991 onward, the measures of democracy refer to the Russian Federation.

function indicate the absence of democracy, the other three indicators suggest the existence of some democratic structures. This discrepancy is explainable with the differences in the functional assumptions and the fact that one-party elections were regularly held in the Soviet Union. For example, the additive index assumes that level of democracy is equal to 0 only if all regime characteristics are equal to 0, whereas the multiplicative index exceeds 0 only if all regime characteristics have a positive value. Since electoral participation was relatively high in the Soviet Union and electoral competition was completely absent (Nohlen and Stöver, 2010), we find that the additive index indicates a much higher level of democracy for the Soviet Union than the multiplicative index. As a consequence, the change of the additive index is less pronounced after the collapse of the Soviet Union.[24]

Our second example is Switzerland, a country that is widely acknowledged for its well established democratic institutions (Nohlen and Stöver, 2010). Our democracy indices reflect this institutional stability because we do not observe any significant

[24]A similar reasoning applies with respect to the other indices. To avoid redundancies, we only discuss the additive and the multiplicative index.

Figure 3 Democracy in the Switzerland (1919 – 2018).



Notes: The figures show the level of democracy of Switzerland, depending on how we aggregate our ten regime characteristics.

decline in the degree of democratization (see Figure 3). Another fact for which Switzerland is well known is that it was the last European country that introduced female suffrage at the national level. Figure 3 shows that our six indices react differently to enfranchisement of women in 1971: while the machine learning index indicates a significant increases in the level of democracy, the other five indicators change only marginally and still suggest a lack of democracy in Switzerland. The key reason for the reduced levels is the functional assumption that the degree of democratization is equal to 1 only if all regime characteristics reach their highest value.

5 Differences in the numerical form

To study whether the decision on the scale of the democracy index has notable consequences for the results of OLS and 2SLS regressions, we need continuous and dichotomous indicators that are otherwise equivalent. We can address this issue in two ways: first, we can use an extended version of the machine learning approach developed by Gründler and Krieger (2016), or second, we can exploit a continuous index and define a threshold value up to which a regime can be considered as democratic. Since these two approaches produce different results, we discuss them separately.

5.1 Machine Learning Approach

A special feature of the machine learning method is that it does not need manual interventions to produce (otherwise identical) continuous and dichotomous indices. This feature exists because the SVM toolbox includes classification and regression methods that operate in a very similar manner (for details, see Appendix B). We exploit this methodological flexibility to provide novel evidence on the question of how a change from a continuous index to a dichotomous index affects regressions results.

5.1.1 Estimation results

In Columns 1 and 3 of Table 5, we show that the OLS estimate of the effect of democracy on economic growth decreases from 1.7 percent to 1.0 percent when we apply the dichotomous machine learning index rather than the continuous machine learning index. This decline is statistically significant at conventional levels and lends support to Elkins' (2000) results suggesting that continuous indices produce larger estimates than dichotomous measures. In his study, Elkins (2000) does not examine whether the difference in the size of the regression coefficients persists in 2SLS regressions. Columns 2 and 4 fill this gap. Again, we observe that the dichotomous index yields a smaller effect of transitions towards democracy than the continuous index. However, the difference in the effect is not statistically significant (p-value: 0.373). Table 5 thus suggests that the continuous and the dichotomous indicator behave differently in OLS regressions and similarly in 2SLS regressions.

To illustrate the external validity of the results reported in Table 5, we conduct the same robustness checks as in Section 4.1.3. Appendix Tables C.17 – C.23 show their results. Three findings are especially notable: First, the estimate produced by the continuous indicator always exceeds the estimate produced by the dichotomous

Table 5 Consequences of using different numerical forms — Machine learning indices.

	Continuous SVM index		Dichotomous SVM index	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Democracy	0.017^{***}	0.032^{***}	0.010^{***}	0.027^{***}
	(0.003)	(0.006)	(0.002)	(0.005)
Income_{t-1}	1.185^{***}	1.181^{***}	1.187^{***}	1.182^{***}
	(0.045)	(0.045)	(0.045)	(0.045)
Income_{t-2}	-0.108	-0.106	-0.109	-0.106
	(0.070)	(0.070)	(0.070)	(0.070)
Income_{t-3}	-0.084^{**}	-0.084^{**}	-0.085^{**}	-0.084^{**}
	(0.036)	(0.036)	(0.036)	(0.037)
Income_{t-4}	-0.008	-0.009	-0.007	-0.009
	(0.019)	(0.019)	(0.019)	(0.019)
Observations	10026	10026	10026	10026
Countries	163	163	163	163
Equal. (p-val.)	–	–	0.016	0.373
First-stage	–	0.943^{***}	–	0.711^{***}
SW (F-stat.)	–	118.91	–	114.45
AR (p-val.)	–	0.000	–	0.000
Long-run effect	1.133	1.780	0.708	1.563

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

indicator. Second, the difference in the second-stage estimates is never statistically significant at conventional levels. Third, the OLS estimates differ significantly in many but not in all robustness checks.[25]

5.1.2 Explanation

Basic econometric theory implies that classical measurement error biases an OLS estimate towards zero and that the attenuation bias disappears when applying an instrumental variable approach (Angrist and Pischke, 2009). Thus, differences in the extent of classical measurement error can serve as an explanation for the results presented in Section 5.1.1 if the dichotomous index suffers, on average, more from classical measurement error than the continuous index. Since our machine learning approach produces confidence intervals whose lengths can serve as proxies for the extent of classical measurement error, we can examine whether this logic actually

[25]The difference in the OLS estimates is statistically insignificant, especially if we use 5-year data (see Appendix Tables C.17 and C.20). This result is consistent with the explanation that we will provide in the next section because data averaging mitigates the impact of classical measurement error.

applies.

Alvarez et al. (1996) suggest that continuous and dichotomous democracy indices suffer from different types of classical measurement error: continuous indices have many small errors, whereas dichotomous indices have few but large errors. Our machine learning indices fit this pattern because the median/mean length of a 95 percent confidence interval is smaller/larger for the dichotomous index. The larger mean value is also consistent with our presumption that classical measurement error is more frequent in the dichotomous index.

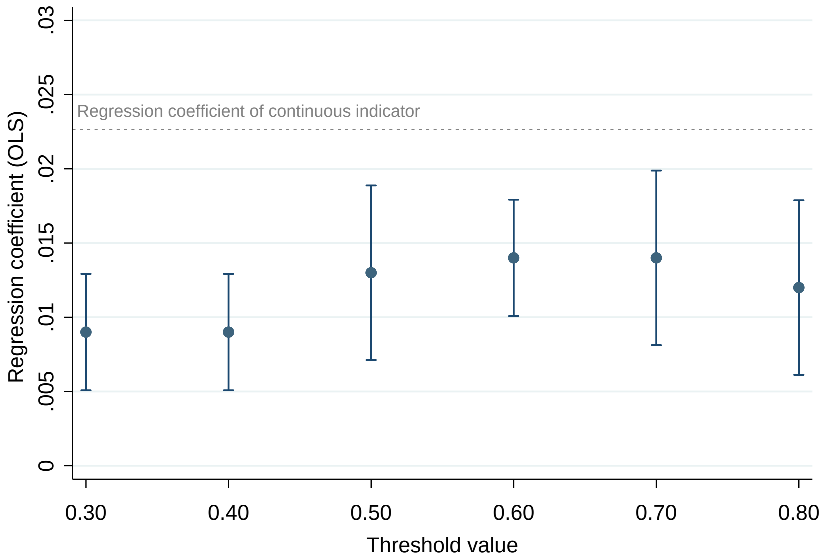
Since our fixed effect model exploits within-country variation in the degree of democratization to identify the effect of democracy on economic growth, classical measurement error poses a problem for the OLS estimator especially if it occurs during political transitions. Therefore, if different attenuation biases explain the differences in the magnitudes of the OLS estimates, we should observe that the dichotomous index has, on average, larger confidence intervals than the continuous index if a regime change takes place. To test whether this prediction holds, we investigate the 95 percent confidence intervals of the country-year observations for which Papaioannou and Siourounis (2008) report a transition. As expected, we find that the dichotomous indicator has larger confidence intervals than the continuous indicator (mean lengths: 0.20 & 0.12).[26] In Appendix Figure C.1, we illustrate the differences in the length of the confidence intervals based on the example of the Russian Federation.

5.2 Defining threshold values

A second way to transform non-binary regime characteristics into a dichotomous measure of democracy is to select a continuous index and to set a threshold up to which a regime can be classified as democratic. Several scholars criticize this procedure especially because the level of such a threshold value is arbitrary (see Bogaards, 2012). From an empirical perspective, this arbitrariness is problematic if the choice of the threshold affects regression results. To examine this concern, we compare the performance of continuous and dichotomous indices and test how the

[26]We observe the same pattern if we use the dichotomous indices of Bjørnskov and Rode (2019) or Boix et al. (2013) to identify regime changes.

Figure 4 Consequences of using different numerical forms — Threshold approach (OLS).



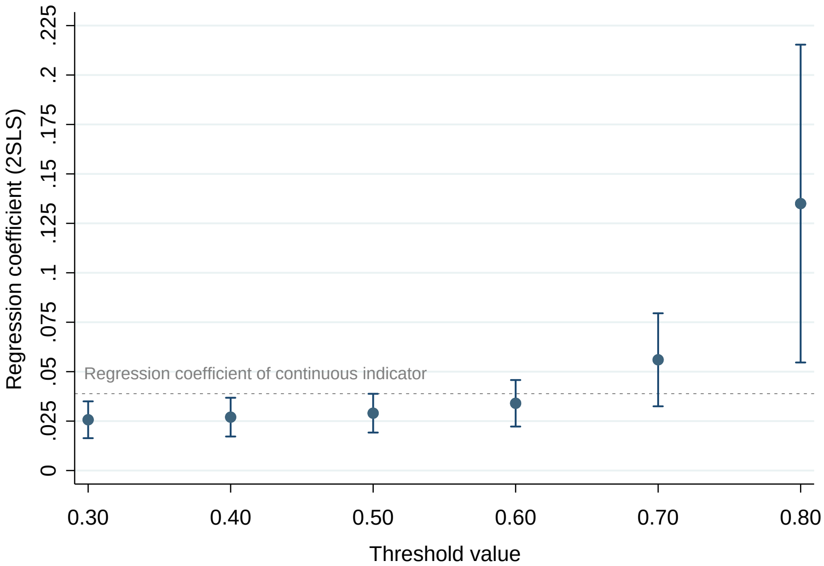
Notes: This figure presents the results of seven OLS regressions (for the tabulated results, see Appendix Figure C.24). The dependent variable is always the log of GDP per capita and all regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the regressions is the democracy index. The dashed line reflects the regression coefficient produced by the multiplicative index. The dots show the estimates produced by the dichotomous indices. At the horizontal axes, we present the threshold that a regime must reach to be labeled as democratic. The vertical lines indicate the 95 percent confidence intervals of the point estimates.

choice of the threshold affects the regression coefficients produced by a dichotomous index.

We use six different threshold values to transform our continuous indices into dichotomous indices and repeat our baseline regressions with each of these binary measures of democracy. Figures 4 and 5 illustrate the results of these regressions when using the multiplicative index (for the tabulated results, see Appendix Tables C.24 and C.25). In both figures, the dashed line reflects the regression coefficient produced by the multiplicative indicator, while the dots show the estimates of the dichotomous indicators. At the horizontal axes, we present the threshold that a regime must reach to be labeled as democratic. The vertical lines indicate the 95 percent confidence intervals associated with the point estimates.

We find that the choice of the threshold value has only a minor impact on the magnitude of the OLS estimates (see Figure 4). We also see that the dichotomous indices create significantly smaller estimates than the continuous index, which is consistent with the pattern that we find for the machine learning indices and the

Figure 5 Consequences of using different numerical forms — Threshold approach (2SLS).



Notes: This figure presents the results of seven OLS regressions (for the tabulated results, see Appendix Figure C.25). The dependent variable is always the log of GDP per capita and all regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. The only difference between the regressions is the democracy index. The dashed line reflects the regression coefficient produced by the multiplicative index. The dots show the estimates produced by the dichotomous indices. At the horizontal axes, we present the threshold that a regime must reach to be labeled as democratic. The vertical lines indicate the 95 percent confidence intervals of the point estimates.

conclusion drawn by Elkins (2000).[27]

The 2SLS estimates differ in two remarkable ways from the OLS estimates (see Figure 5). First, we observe that the size of the estimated effect of democracy on economic development increases when we set a higher threshold. The reason is a weaker first-stage relationship (see Appendix Table C.25). Second, the choice of the threshold value determines whether (and how) the scale of an index affects the regression result. If we define a low threshold, the continuous index produces significantly larger 2SLS estimates than the dichotomous index. For intermediate thresholds, we find no statistically significant difference. If the threshold is high, the estimates of the dichotomous measure significantly exceed the estimates of the continuous measure.

Appendix Tables C.26 and C.27 suggest that we obtain similar results when we

[27]As in Section 5.1, we think that differences in the extent of classical measurement error explain the differences in the OLS estimates produced by the continuous and dichotomous indices. Since no confidence intervals exist for the dichotomous indidces, we cannot investigate whether this explanation actually applies.

replace the multiplicative indicator with other indicators. The only small change relates to the threshold at which the continuous and dichotomous index produce similar 2SLS estimates since we observe that this threshold is not the same for all aggregation methods. Our results also hold when we repeat the robustness checks that we proposed in Section 4.1.3 and Section 5.1.1 (see Appendix Tables C.28 – C.41).

Since arbitrary decisions should not significantly affect the results of empirical studies, we believe that the concerns of those social scientists who criticized the dichotomization of continuous indices are legitimate (see Bogaards, 2012, Cheibub et al., 2010). We also share the view that only "original" dichotomous indicators should be used in regression analyses.[28]

6 Conclusion

In economics and other social sciences, building indices is the conventional way of summarizing multidimensional data. A key challenge when designing indicators is to address the question of how to transform the raw data into an uni-dimensional measure. The literature offers various aggregation methods, but does not provide information on the consequences arising from the decisions that scholars make when choosing a specific method. We therefore investigate whether democracy indicators that differ with regard to their numerical form and aggregation function produce different results in regression analyses. Our results suggest that both the choice between a continuous and dichotomous scale and the choice of a functional form affects the results of OLS and 2SLS regressions in a statistically significant manner. We also indicate reasons for why these decisions have notable empirical consequences.

Our findings give rise to the following recommendations. First, when producing an index, scholars should check whether their aggregation tool agrees with their conceptual assumptions. A mismatch exists, for example, if an additive aggregation function is used for a concept that treats each aspect as a necessary condition.

[28] Examples include the dichotomous machine learning index (see Section 5.1) and the index of Boix et al. (2013). The dichotomous index of Bjørnskov and Rode (2019) should be used with caution since it systematically underestimates the consequences of political transition due to conceptual issues (for details, see Knutsen and Wig, 2015).

Second, scholars should provide some example that illustrate how their indicator behaves, especially at the upper and lower end of the spectrum. Third, robustness tests are indispensable: the reader needs to know how the estimates react if the preferred aggregation method is replaced by alternative methods. This transparency is especially important when effect sizes form the basis for policy recommendations. Fourth, dichotomizing continuous indices is problematic because the thresholds are arbitrary and influence regression results. Finally, all indices suffer from classical measurement errors. Establishing causality thus needs identification strategies that seriously address this issue. If such a regression tool does not exist and a choice must be made between a continuous and a dichotomous index, we propose to use continuous indices since they are, on average, less prone to classical measurement errors than dichotomous indices.

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. (2019). Democracy does cause growth. *Journal of Political Economy*, 127(1):47–100.

African Election Database (2019). A database of election results in Sub-Saharan Africa. `http://africanelections.tripod.com/`.

Aidt, T. S. and Franck, R. (2015). Democratization under the threat of revolution: Evidence from the Great Reform Act of 1832. *Econometrica*, 83(2):505–547.

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.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's* *companion*. Princeton University Press.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104(1):184–198.

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. (2019). Regime types and regime change: A new dataset on democracy, coups, and political institutions. *Review of International Organizations*.

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

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.

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review*, 45(3):370–390.

Bollen, K. A. (1990). Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development*, 25(1):7–24.

Bolt, J., Inklaar, R., de Jong, H., and van Zanden, J. L. (2018). Rebasing 'maddison': New income comparisons and the shape of long-run economic development. *GGDC Research Memorandum*.

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.

Brecke, P. (1999). Violent conflicts 1400 ad to the present in different regions of the world.

Breiman, L. et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231.

Brückner, M. and Ciccone, A. (2011). Rain and the democratic window of opportunity. *Econometrica*, 79(3):923–947.

Carr, A. (2019). Psephos: Adam carr's electoral archive. `http://psephos.adam-carr.net/`.

Casper, G. and Tufis, C. (2003). Correlation versus interchangeability: The limited robustness of empirical findings on democracy using highly correlated data sets. *Political Analysis*, 11(2):196–203.

Cervellati, M., Jung, F., Sunde, U., and Vischer, T. (2014). Income and democracy: Comment. *American Economic Review*, 104(2):707–719.

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

Coppedge, M., Alvarez, A., and Maldonado, C. (2008). Two persistent dimensions of democracy: Contestation and inclusiveness. *Journal of Politics*, 70(03):632–647.

Coppedge, M., Gerring, J., Altman, D., Bernhard, M., Fish, S., Hicken, A., Kroenig, M., Lindberg, S. I., McMann, K., Paxton, P., et al. (2011). Conceptualizing and measuring democracy: A new approach. *Perspectives on Politics*, 9(2):247–267.

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

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.

Diamond, L. J. (2002). Thinking about hybrid regimes. *Journal of Democracy*, 13(2):21–35.

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

Doucouliagos, H. and Ulubaşoğlu, M. A. (2008). Democracy and economic growth: A meta-analysis. *American Journal of Political Science*, 52(1):61–83.

Dreher, A. (2006). Does globalization affect growth? Evidence from a new index of globalization. *Applied Economics*, 38(10):1091–1110.

Durlauf, S. N., Johnson, P. A., and Temple, Jonathan R. W. (2005). Growth econometrics. In *Handbook of Economic Growth*. Elsevier.

Elkins, Z. (2000). Gradations of democracy? Empirical tests of alternative conceptualizations. *American Journal of Political Science*, pages 293–300.

Flachaire, E., García-Peñalosa, C., and Konte, M. (2014). Political versus economic institutions in the growth process. *Journal of Comparative Economics*, 42(1):212–229.

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

Goertz, G. (2006). *Social science concepts: A user's guide*. Princeton University Press.

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

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

Gundlach, E. and Paldam, M. (2009). A farewell to critical junctures: Sorting out long-run causality of income and democracy. *European Journal of Political Economy*, 25(3):340–354.

Gygli, S., Haelg, F., Potrafke, N., and Sturm, J.-E. (2019). The KOF globalisation index–revisited. *Review of International Organizations*.

Hale, H. E., McFaul, M., and Colton, T. J. (2004). Putin and the "delegative democracy" trap: Evidence from Russia's 2003-04 elections. *Post-Soviet Affairs*, 20(4):285–319.

Harding, R. and Stasavage, D. (2013). What democracy does (and doesn't do) for basic services: School fees, school inputs, and african elections. *Journal of Politics*, 76(1):229–245.

Hawken, A. and Munck, G. L. (2013). Cross-national indices with gender-differentiated data: What do they measure? How valid are they? *Social Indicators Research*, 111(3):801–838.

Huntington, S. P. (1993). *The third wave: Democratization in the late twentieth century*. University of Oklahoma Press.

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

International Foundation for Electoral Systems (2019). Election database. `http://www.ifes.org/`.

International Institute for Democracy and Electoral Assistance (2019). Voter turnout database. `http://www.idea.int/`.

Knutsen, C. H. (2015). Why democracies outgrow autocracies in the long run: Civil liberties, information flows and technological change. *Kyklos*, 68(3):357–384.

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.

Krieckhaus, J. (2004). The regime debate revisted: A sensitivity analysis of democracy's economic effect. *British Journal of Political Science*, 34(04):635–655.

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.

Lipset, S. M. (1959). Some social requisites of democracy. *American Political Science Review*, 53(1):69–105.

Madsen, J. B., Raschky, P. A., and Skalil, A. (2015). Does democracy drive income in the world, 1500–2000? *European Economic Review*, 78(1):175–195.

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.

Merkel, W. (2004). Embedded and defective democracies. *Democratization*, 11(5):33–58.

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

Murtin, F. and Wacziarg, R. (2014). The democratic transition. *Journal of Economic Growth*, 19(2):141–181.

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.

Olson, M. (1993). Dictatorship, democracy, and development. *American Political Science Review*, 87(3):567–576.

Papaioannou, E. and Siourounis, G. (2008). Democratisation and Growth. *Economic*

Journal, 118(532):1520–1551.

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.

Persson, T. and Tabellini, G. (2005). *The economic effects of constitutions*. MIT Press.

Persson, T. and Tabellini, G. (2006). Democracy and development: The devil in the details. *American Economic Review*, 96(2):319–324.

Persson, T. and Tabellini, G. (2008). The growth effect of democracy: Is it heterogenous and how can it be estimated? In *Institutions ans Economic Performance*. Harvard University Press.

Persson, T. and Tabellini, G. (2009). Democratic Capital: The Nexus of Political and Economic Change. *American Economic Journal: Macroeconomics*, 1(2):88–126.

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

Przeworski, A. (2000). *Democracy and development: Political institutions and well-being in the world, 1950-1990*. Cambridge University Press.

Rodrik, D. and Wacziarg, R. (2005). Do democratic transitions produce bad rconomic outcomes? *American Economic Review*, 95(2):50–55.

Sakwa, R. (2005). *The rise and fall of the Soviet Union*. Routledge.

Sakwa, R. (2010). *The crisis of Russian democracy: The dual state, factionalism and the Medvedev succession*. Cambridge University Press.

Sanderson, E. and Windmeijer, F. (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.

Schedler, A. (2002). The menu of manipulation. *Journal of Democracy*, 13(2):36–50.

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.

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

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

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

(US).

Tavares, J. and Wacziarg, R. (2001). How democracy affects growth. *European Economic Review*, 43(8):1341–1378.

Teorell, J. (2010). *Determinants of democratization: Explaining regime change in the world, 1972–2006*. Cambridge University Press.

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.

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

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

Vapnik, V. and Chervonenkis, A. (1964). A note on one class of the algorithms of pattern recognition. *Automation and Remote Control*, 25(6):821–837.

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

Wintrobe, R. (1990). The tinpot and the totalitarian: An economic theory of dictatorship. *American Political Science Review*, 84(3):849–872.

For online publication

A Further information on the Machine Learning indices

The machine learning tool that we developed in an earlier work (see Gründler and Krieger, 2016) needs observations—called *priming data*—which the SVM techniques can exploit to "learn" the functional relationship between the regime characteristics and the degree of democratization. To give our machine learning approach the opportunity to produce plausible measures of democracy, the priming data must meet two conditions: first, the country-year observations belonging to the priming data must be correctly labeled, and second, these observations need to reflect the institutional heterogeneity among autocracies and democracies. In this section, we provide evidence suggesting that both conditions hold.

A.1 Adequacy of the priming data

Our priming data includes 2.728 country-year observations: 1.308 democracies and 1.420 autocracies (for a list, see Appendix Table A.4). We used the handbooks published by Dieter Nohlen (1999, 2001, 2005, 2010) and many other sources to examine whether the selected observations are correctly labeled. We detected only four controversial observations: The first two are the country-years for Switzerland in 1952/1970 because women could not vote in national elections. The third is the observation for Israel in 1999 because Palestinians living in the Gaza Strip and the West Bank were not allowed to participate in elections of the Israeli parliament. The last case is the observation for Uganda in 1980, which is labeled as autocratic. This label is reasonable given that a coup occurred in May 1980. However, in December 1980, a multiparty election was hold, which speaks against the adequacy of this label.[1]

We also investigated how other democracy indicators evaluate the country-year observations included in the priming data. Table A.1 shows the results of our tests and suggests that a consensus exists on the selected regimes. For example, we find that 99.84 percent of the democratically labeled observations belong to the highest

[1] None of our regression results change in a notable manner if we exclude these four observations from the priming data.

category of the Lexical Index of Electoral Democracy (LIED) developed by Skaaning et al. (2015).

Table A.1 Agreement among democracy indices about labeled observations.

	Democratic regimes				
	Polity (≥ 7)	LIED ($= 6$)	ANRR ($= 1$)	BMR ($= 1$)	BR ($= 1$)
Overlap	0.9943	0.9984	1.0000	1.0000	1.0000
	Autocratic regimes				
	Polity (≤ -7)	LIED (≤ 1)	ANRR ($= 0$)	BMR ($= 0$)	BR ($= 0$)
Overlap	0.9511	0.9441	0.9993	0.9993	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), and the dichotomous indices by Acemoglu et al. (2019), Boix et al. (2013), and Bjørnskov and Rode (2019).

A.2 Representativeness of the priming data

Autocratic and democratic regimes take different forms. The priming data need to reflect this institutional heterogeneity, since otherwise we may produce indices that discriminate in favor of specific regime types. Below, we allay this concern with a number of tests.

Cheibub et al. (2010) distinguish between four different types of dictatorship: civil dictatorship, communist dictatorship, military dictatorship, and royal dictatorship. The first column of Table A.2 shows that our priming data includes autocracies of each of these four types.

Schedler (2002) discerns between electoral autocracies and closed autocracies: the first category includes autocracies in which non-competitive elections take place, while the second category consists of regimes in which the government does not organize national elections. The second and third column of Table A.2 illustrate that our priming data includes both electoral and closed autocracies.

The theories developed by Olson (1993) and Wintrobe (1990) suggest that the durability of a regime can serve as a classification criterion for autocracies. We distinguish between three types: (i) autocracies that exist for less than six years, (ii) autocracies that exist for more than 6 years and less than 26 years, and (iii) autocracies that exist for at least 26 years. The last three columns of Table A.2 indicate that the durations of the autocratic regimes included in the priming data differ considerably.

Table A.2 Representativeness of autocratic regimes in priming data.

	Regime	**Elections**		**Duration**		
	Σ	*Yes*	*No*	*≤ 5 years*	*6 – 25 years*	*> 25 years*
Civil dictator.	199	141	58	64	115	20
Communist dictator.	386	269	117	49	153	184
Military dictator.	429	119	310	154	229	46
Royal dictator.	406	84	322	37	128	241
	1,420	*613*	*807*	*304*	*625*	*491*

Notes: Our machine learning approach requires that the priming data reflects the institutional heterogeneity among autocracies and democracies. This table presents how the autocratically labeled country-year observations are distributed over different categories. The data comes from Bjørnskov and Rode (2019) and Geddes et al. (2014).

Cheibub et al. (2010) distinguish democratic regimes according to the balance of power between the parliament and the government. Their classification consists of three categories: parliamentary regimes, semi-presidential regimes, and presidential regimes. The first column of Table A.3 suggests that our priming data includes democratic regimes of all three types.

Persson and Tabellini (2005) state that legislative authorities are elected through different voting procedures (proportional vs. majoritarian systems) and have different structures (unicameral vs. bicameral systems). Columns 2 – 5 of Table A.3 report that the priming data reflects these institutional heterogeneities.

Table A.3 Priming data — Representativeness of democratic regimes.

	Regime	**Legislature**		**Voting system**	
	Σ	*Unicameral*	*Bicameral*	*Proportional*	*Majoritarian*
Parliamentary demo.	792	366	426	606	186
Semi-presidential demo.	306	102	204	260	46
Presidential demo.	210	63	147	173	37
	1,308	*531*	*777*	*1,039*	*269*

Notes: Our machine learning approach requires that the priming data reflects the institutional heterogeneity among autocracies and democracies. This table presents how the democratically labeled country-year observations are distributed over different categories. The data comes from Bjørnskov and Rode (2019).

Table A.4 Priming data — Selected country-years.

Country	Observations	Years
	Democratic regimes (1308 observations)	
Australia	51	1964, 1966–67, 1969–2016
Austria	63	1950–59, 1961, 1963–65, 1967–2015
Barbados	2	1981–82
Belgium	55	1955, 1963–2016
Brazil	19	1995–98, 2000–14
Canada	42	1972–81, 1983–2013, 2016
Chile	15	2002–16
Costa Rica	41	1975–80, 1982–2016
Cyprus	22	1989–2010
Czech Republic	23	1991–2013
Denmark	67	1950–2016
Estonia	24	1993–2016
Finland	35	1967–1971, 1987–2016
France	47	1970–2016
Germany	48	1950–52, 1972–2016
Greece	28	1984–2011
Hungary	8	1994–1997, 2004–05, 2007, 2009
Iceland	47	1967–2013
Ireland	40	1968, 1976–81, 1983–2015
Israel	1	1999
Italy	48	1960–61, 1963–1976, 1983–1992, 1994–2012, 2014–16
Japan	14	1980–90, 2010–14
Latvia	2	2013–14
Luxembourg	60	1951–53, 1960–2016
Malta	27	1973–75, 1989–2012
Mauritius	1	2004
Netherlands	65	1952–2016
New Zealand	54	1963–2016
Norway	49	1968–2016
Poland	24	1991–2014
Portugal	33	1984–2016
Slovakia	12	1999–2012
Slovenia	12	1997, 2002–11, 2013–15
Spain	29	1980–82, 1987–2012
St. Kitts and Nevis	1	2010
Sweden	50	1967–2016
Switzerland	48	1952, 1970–2016
Taiwan	1	2005
United Kingdom	35	1961, 1964, 1975–89, 1999–2016
United States	37	1968–69, 1976–77, 1983–86, 1988–2016
Uruguay	28	1989–2016
	Autocratic regimes (1420 observations)	
Afghanistan	34	1950–63, 1978–91, 1996–2001
Albania	29	1951–53, 1955, 1957–61, 1965, 1970, 1981–89
Algeria	14	1965–76, 1985, 1994
Angola	18	1975–92
Argentina	12	1966–71, 1977–82
Bahrain	28	1971–73, 1976–2000
Benin	10	1974–79, 1984, 1986–87, 1989
Bhutan	57	1950–2006
Bolivia	7	1972–77, 1980
Brazil	2	1968–69
Bulgaria	8	1980, 1983–89
Burkina Faso	1	1965
Burma (Myanmar)	39	1963–1974, 1983–2008, 2010
Burundi	16	1967–81, 1988
Cambodia	16	1953, 1966–67, 1979–91
Cameroon	1	1988
Central African Rep.	15	1967–1980, 1988
Chad	17	1962, 1970, 1972–74, 1976–77, 1980–89
Chile	14	1975–88
China	44	1950–78, 1985, 1989–97, 2000, 2013–16
Democratic Rep. of Congo	16	1965–76, 1983–87, 1989
Rep. of Congo	7	1968–1972, 1978–79
Cuba	48	1958, 1960–2005, 2007
Czech Republic	5	1950–1954
Dominican Rep.	10	1950–1953, 1955–1960
⋮	⋮	⋮

Table A.4 Priming data — Selected country-years (continued).

Country	Observations	Years
⋮	⋮	⋮
Egypt	3	1954–56
Equatorial Guinea	17	1973–82, 1985–91
Eritrea	15	2002–16
Ethiopia	32	1950–59, 1961–62, 1965, 1970–72, 1974–86, 1988–90
Gabon	1	1968
Germany (East)	25	1953, 1960–66, 1969, 1972–88
Ghana	1	1965
Greece	6	1968–73
Guatemala	3	1956, 1964–54
Guinea	18	1958–60, 1967–68, 1972–84
Guinea-Bissau	2	1980, 1985
Haiti	20	1950, 1963–64, 1967, 1971–84, 1992–93
Indonesia	1	1965
Iran	19	1953–64, 1966–1974, 1976
Iraq	40	1963–2002
Ivory Coast	1	1966
Jordan	22	1950, 1957–59, 1961–64, 1967–68, 1970, 1974–83
Korea (North)	59	1958–2016
Kuwait	10	1965–66, 1976, 1979–80, 1986–1990
Laos	40	1976–89, 1991–2016
Lesotho	2	1970–71
Liberia	2	1980–81
Libya	45	1952–55, 1970–2010
Malawi	22	1964, 1966, 1972–87, 1989–92
Maldives	4	1965–68
Mali	7	1975–78, 1980–81, 1985
Mauritania	4	1979–82
Mongolia	9	1950–51, 1980, 1983–88
Morocco	12	1956–1962, 1965–69
Mozambique	15	1976–1990
Nepal	10	1950–51, 1960–65, 1967–68
Niger	4	1979–82
Nigeria	1	1966
Oman	31	1970–2000
Pakistan	1	1980
Panama	3	1969–71
Paraguay	5	1954, 1956–57, 1960–61
Peru	1	1969
Philippines	5	1973–77
Portugal	16	1951–54, 1956–59, 1961–64, 1966–68, 1970
Qatar	47	1970–2016
Romania	6	1983–88
Russia (Soviet Union)	9	1950–52, 1967–69, 1984-86
Rwanda	5	1974–1977, 1980
Saudi Arabia	67	1950–2016
Serbia (Yugoslavia)	10	1950, 1954–57, 1975–79
Somalia	19	1970–80, 1983–90
Spain	18	1950–67
Sudan	16	1959–63, 1989–97, 1999, 2001
Swaziland	11	1974–77, 1984, 1986–87, 1989–92
Syria	28	1961, 1965, 1970–73, 1983–2002, 2011–12
Taiwan	20	1950–69
Togo	14	1967–79, 1985
Tunisia	5	1956–1959, 1962
Turkmenistan	20	1992–2012
Uganda	9	1972–1980
United Arab Emirates	41	1971–2011
Uruguay	3	1976–78
Uzbekistan	16	1995–06, 2008, 2010–12
Venezuela	3	1954–56
Vietnam	20	1954–59, 1985–86, 1989–2000
Yemen (Yemen North)	31	1950–70, 1978–1987

Notes: This table reports the country-year observations that are part of the priming data. The selection is based on the indicies developed by Pemstein et al. (2010) and Teorell et al. (2019).

B Support Vector Machines

Support Vector Machines (SVM) is a frequently used machine learning technique designed for pattern recognition. SVM aims at revealing an unknown functional relationship $\mathfrak{F}\colon \mathcal{X} \rightarrow \mathcal{Y}$ that links a set of input characteristics $\mathbf{x} = (x_1, \ldots, x_m) \in \mathcal{X} \subseteq \mathbb{R}^m$ to an outcome variable $y \in \mathcal{Y}$ for all observations in the sample $\mathcal{S} = \{(\mathbf{x}_i, y_i) \,|\, i = 1, \ldots, n\}$:

$$\mathfrak{F}(\mathbf{x}_i) \stackrel{!}{=} y_i \qquad \forall\, i = 1, ..., n. \tag{23}$$

In contrast to conventional tools of statistical modeling—such as Ordinary Least Squares or Generalized Methods of Moments—machine learning techniques do not need prior assumptions about the shape of the functional relationship between the input characteristics and the outcome variable. They rather learn without being explicitly programmed (Breiman et al., 2001). The related literature distinguishes between supervised and unsupervised machine learning techniques.[2] SVM belongs to the former type since its application requires observations for learning the rule that maps the inputs onto the output (Steinwart and Christmann, 2008).[3]

The mathematical foundations of SVM methods and their properties with regard to prediction accuracy, statistical robustness, and practicability are well documented (see Abe, 2005, Bennett and Campbell, 2000, Guenther and Schonlau, 2016, Steinwart and Christmann, 2008). In this study, we use two common SVM methods to arrive at dichotomous classifications and to run non-linear regressions. In this section, we present the mathematical formulations of the Support Vector Classification and the Support Vector Regressions.[4]

[2]The application of supervised machine learning methods requires the existence of observable input variables ($\mathbf{x}$) and an observable output variable (y). The main objective is to estimate a mapping function that allows predicting the output variable for new input data. In contrast, unsupervised machine learning techniques are applied when no output variable exists and the available data need to be structured.

[3]In this context, "learning the rule" means that an empirical model is estimated which adequately predicts the output y of any input $\mathbf{x}$; it does not mean that SVM provides a closed form description of the functional relationship that facilitates a causal interpretation of the impact of the input characteristic x_j ($j = 1, \ldots, m$) on the outcome y.

[4]For further reading, we refer interested readers to the works of Abe (2005), Smola and Schölkopf (2004), Steinwart and Christmann (2008) and Vapnik (1995, 1998)

B.1 Support Vector Classification

The Support Vector Classification (SVC) is a non- linear extension of the General Portrait Algorithm (GPA) developed by Vapnik and Lerner (1963) and Vapnik and Chervonenkis (1964). In its initial form, the GPA assumes the existence of some hyperplanes:

$$E_{\mathbf{w},b}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \qquad \mathbf{w} \in \mathbb{R}^m \,,\ ||\mathbf{w}|| = 1 \,,\ b \in \mathbb{R} \,,\ \mathbf{x} \in \mathbb{R}^m \tag{24}$$

that can separate the observations in the sample $\mathcal{S}$ according to their labels $y \in \{-1, 1\}$.[5] Graph (I) in Figure B.1 illustrates this separation in a one-dimensional example.

The primary objective of the GPA is to find a linear classification function that assigns any input $\mathbf{x}_i$ to its output z_i $(i = 1, \ldots, n)$. The second Graph in Figure B.1 illustrates that the number of eligible decision functions might be infinite. To arrive at an unique solution, the distance—called the *margin*—between a separating hyperplane and the nearest observation is calculated. GPA selects the hyperplane with the greatest margin in $\mathcal{S}$ (Abe, 2005, Steinwart and Christmann, 2008). Graphs (III) and (IV) in Figure B.1 illustrate this procedure.

In formal terms, the GPA solves the quadratic optimization problem:

$$\min_{\mathbf{w},\, b} \frac{1}{2} \langle \mathbf{w},\, \mathbf{w} \rangle \qquad \text{s.t.} \qquad y_i \cdot (\langle \mathbf{w},\, \mathbf{x}_i \rangle \,+\, b) \,\geq\, 1 \tag{25}$$

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

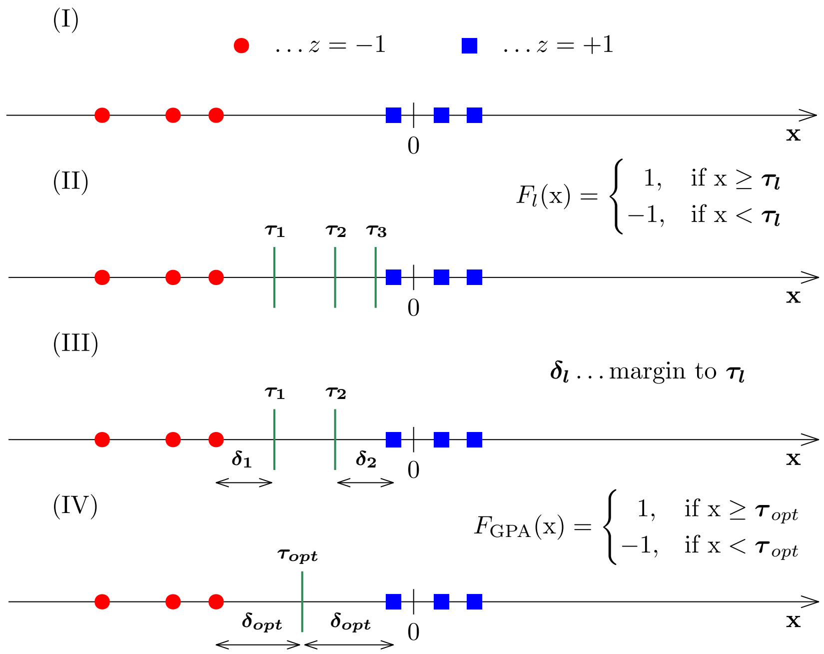
$$\mathfrak{F}(\mathbf{x}) \,=\, \mathrm{sign}\,(\langle \mathbf{w}^*, \mathbf{x} \rangle + b^*) \qquad \text{where} \qquad \mathbf{w}^* \,\in\, \mathbb{R}^m \quad \text{and} \quad b^* \,\in\, \mathbb{R}. \tag{26}$$

The GPA attracts little attention in applied research because a linear separation usually does not exist (see Graph (I) in Figure B.2). Boser et al. (1992) extend the GPA to allow for the estimation of non-linear classification functions. They propose the use of a non-linear function $\Phi\colon \mathcal{X} \to \mathcal{H}$ that maps the input characteristics $\mathbf{x} \in \mathcal{X}$ onto a *Reproducing Hilbert Space* $\mathcal{H}$.[6] The GPA is then applied to the adjusted

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

[6]The non-linear extension suggested by Boser et al. (1992) is based on mathematical theorems that

Figure B.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 δ is calculated. Graph IV illustrates that the GPA selects the hyperplane with the largest margin.

sample $\mathcal{S}_{\mathcal{H}} = \{(\Phi(\mathbf{x}_i), z_i) \,|\, i = 1, \ldots, n\}$ and a dividing hyperplane is computed in $\mathcal{H}$:

$$E^{\mathcal{H}}_{\mathbf{w}^*_{\mathcal{H}}, b^*_{\mathcal{H}}}(\Phi(\mathbf{x})) = \langle \mathbf{w}^*_{\mathcal{H}}, \Phi(\mathbf{x}) \rangle + b^*_{\mathcal{H}} \quad \text{with} \quad \mathbf{w}^*_{\mathcal{H}} \in \mathcal{H} \quad \text{and} \quad b^*_{\mathcal{H}} \in \mathbb{R}. \tag{27}$$

The resultant classification function

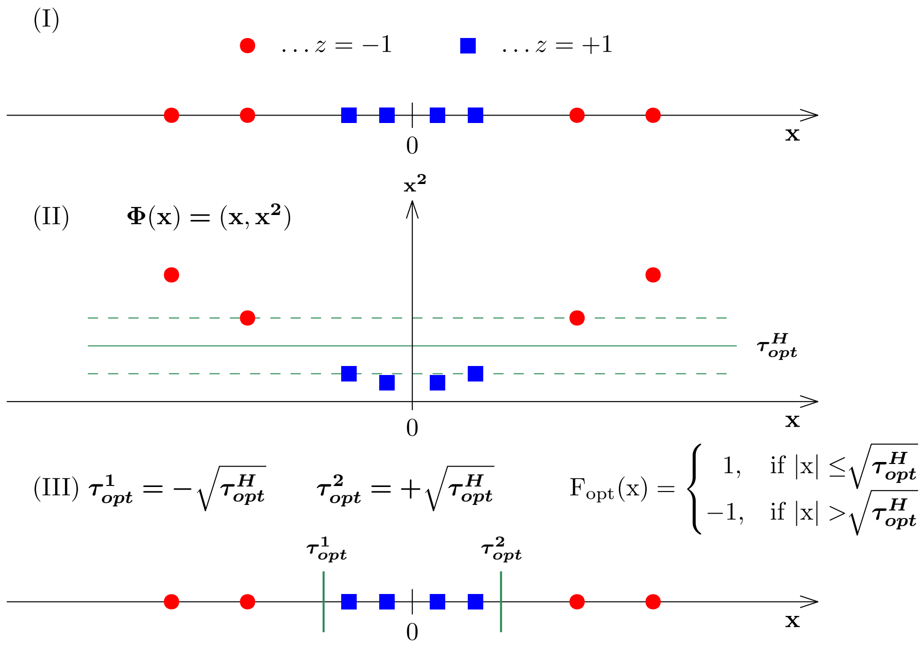
$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}(\langle \mathbf{w}^*_{\mathcal{H}}, \Phi(\mathbf{x}) \rangle + b^*_{\mathcal{H}}) \quad \text{with} \quad \mathbf{w}^*_{\mathcal{H}} \in \mathcal{H} \quad \text{and} \quad b^*_{\mathcal{H}} \in \mathbb{R} \tag{28}$$

is non-linear in $\mathbf{x} \in \mathcal{X}$ (Abe, 2005, Steinwart and Christmann, 2008). Graphs (II) and (III) in Figure B.2 illustrate the mapping approach with the help of a simple example.

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 including slack variables $\xi_i \geq 0$. Together with the non-linear GPA extension of

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 B.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 $\Phi(x) = \left(x, x^2\right)$ 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 $\mathcal{X}$.

Boser et al. (1992), this adjustment yields the optimization problem:

$$\min_{\mathbf{w}_\mathcal{H},\, b_\mathcal{H},\, \xi} \; \frac{1}{2}\langle \mathbf{w}_\mathcal{H},\, \mathbf{w}_\mathcal{H}\rangle \,+\, C \cdot \sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad y_i \cdot (\langle \mathbf{w}_\mathcal{H},\, \Phi(\mathbf{x}_i)\rangle \,+\, b_\mathcal{H}) \,\geq\, 1 \,-\, \xi_i \quad \forall\, i \tag{29}$$

where C denotes a fixed cost parameter for penalizing misclassifications.

If the dimension of $\mathcal{H}$ is large, solving this optimization problem may turn out to be computationally infeasible (Steinwart and Christmann, 2008). In this case, the corresponding dual program:

$$\max_{\alpha \,\in\, [0,C]^n} \; \sum_{i=1}^{n} \alpha_i \,-\, \frac{1}{2} \cdot \sum_{i,j=1}^{n} z_i \cdot z_j \cdot \alpha_i \cdot \alpha_j \cdot \langle \Phi(\mathbf{x}_i),\, \Phi(\mathbf{x}_j)\rangle_\mathcal{H} \quad \text{s.t.} \quad \sum_{i=1}^{n} z_i \cdot \alpha_i \,=\, 0 \tag{30}$$

can be considered where $\alpha_1, \ldots, \alpha_n$ denote the Lagrange multipliers of the primal program. The dual program implies a closed form solution for the classification function:

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n} z_i \alpha_i^* \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x})\rangle_\mathcal{H} \,+\, b_\mathcal{H}^*\right). \tag{31}$$

Because an appropriate *feature map* $\Phi\colon \mathcal{X} \,\rightarrow\, \mathcal{H}$ is usually not known, Schölkopf et al. (1998) apply the "kernel trick", i.e. they replace the unknown inner product $\langle \Phi(\mathbf{x}_i),\, \Phi(\mathbf{x}) \rangle_{\mathcal{H}}$ with a known *kernel function* $\mathfrak{K}\colon\, \mathcal{X} \times \mathcal{X} \to \mathbb{R}$:

$$\mathfrak{F}(\mathbf{x}) \;=\; \operatorname{sign}\left(\sum_{i=1}^{n} z_i \cdot \alpha_i^* \cdot \mathfrak{K}(\mathbf{x}_i,\, \mathbf{x}) \;+\; b_{\mathcal{H}}^* \right).^{7} \tag{32}$$

An observation is called *Support Vector* if its Lagrange multiplier α_i^* is nonzero. The algorithm takes its name from these data points because only Support Vectors influence the shape of the classification function (Abe, 2005, Steinwart and Christmann, 2008).

B.2 Support Vector Regression

In their traditional form, GPA and SVC are limited to applications in which the output (y) comes from a countably finite set. Vapnik (1995, 1998) overcomes this constraint by introducing a method that estimates real-valued functions. The key objective of Support Vector Regression (SVR) 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 labels for all observations in the sample $\mathcal{S} = \{(\mathbf{x}_i,\, y_i) \,|\, i = 1,\, \ldots,\, n\}$:

$$|\, \mathfrak{F}(\mathbf{x}_i) \,-\, y_i \,| \;\overset{!}{\leq}\; \varepsilon \qquad \forall\, i = 1,\, \ldots,\, n. \tag{33}$$

Consider first the case where the regression function is a hyperplane:

$$\mathfrak{F}(\mathbf{x}) \;=\; \langle \mathbf{w},\, \mathbf{x} \rangle \,+\, b \qquad \text{with} \qquad \mathbf{w} \,\in\, \mathbb{R}^m \quad \text{and} \quad b \,\in\, \mathbb{R} \tag{34}$$

and the norm of the slope $\mathbf{w}$ must be minimized. In formal terms, one solves the quadratic optimization problem:

$$\min_{\mathbf{w},\, b}\; \frac{1}{2} \cdot ||\, w \,||^2 \qquad \text{s.t.} \qquad \begin{cases} z_i \,-\, \langle \mathbf{w},\, \mathbf{x}_i \rangle \,-\, b \,\leq\, \varepsilon \quad \forall\, i \\[2ex] \langle \mathbf{w},\, \mathbf{x}_i \rangle \,+\, b \,-\, z_i \,\leq\, \varepsilon \quad \forall\, i. \end{cases} \tag{35}$$

[7] The idea of Schölkopf et al. (1998) is based on a theorem of Mercer (1909), who proves that each kernel function $\mathfrak{K}\colon\, \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is related to a Reproducing Hilbert Space $\mathcal{H}$ with

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

and uses the solution $(\mathbf{w}^*, b^*)$ to specify the regression line.

Since solving the constrained optimization problem (35) often turns out to be impossible, the applicability of a linear SVR is limited. Vapnik (1995, 1998) thus proposes—in a manner similar to SVC—the application of slack variables $(\xi_i^+, \xi_i^-) \in \mathbb{R}_+^2$ $(i = 1, \ldots, n)$ that relax the auxiliary conditions and the use of a feature map $\Phi\colon \mathcal{X} \to \mathcal{H}$ that allows for non-linear estimations:

$$\min_{\mathbf{w}, b, \xi_i^+, \xi_i^-} \frac{1}{2} \| w \|^2 + C \cdot \sum_{i=1}^{n} \left(\xi_i^+ + \xi_i^-\right) \quad \text{s.t.} \quad \begin{cases} z_i - \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^+ \\ \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b - z_i \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0. \end{cases} \tag{36}$$

To avoid computational issues when the dimension of $\mathcal{H}$ is large, the corresponding dual problem

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

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

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

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n} \left(\alpha_i^+ - \alpha_i^-\right) \cdot \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}} + b_{\mathcal{H}}^*. \tag{37}$$

Since the mapping function $\Phi\colon \mathcal{X} \to \mathcal{H}$ is still not known, the kernel trick can again be applied to replace the unknown inner product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}}$ with a kernel $\mathfrak{K}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The shape of the non-linear regression function

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \cdot \mathfrak{K}(\mathbf{x}_i, \mathbf{x}) + b_{\mathcal{H}}^*. \tag{38}$$

depends only on observations—called *Support Vectors*—whose Lagrangian multipliers (α_i, α_i^*) are different from zero (Smola and Schölkopf, 2004).

C Additional Tables and Figures

Table C.1 Creation of ordinal regime characteristics.

Regime Characteristic	Category	Range
Suffrage	0	$x = 0.0$
	1	$x \in (0.0, 0.1]$
	2	$x \in (0.1, 0.3]$
	3	$x \in (0.3, 0.5]$
	4	$x \in (0.5, 1.0]$
Voter-Population ratio	0	$x = 0.0$
	1	$x \in (0.0, 0.1]$
	2	$x \in (0.1, 0.3]$
	3	$x \in (0.3, 0.5]$
	4	$x \in (0.5, 1.0]$
Voter Turnout	0	$x = 0.0$
	1	$x \in (0.0, 0.1]$
	2	$x \in (0.1, 0.3]$
	3	$x \in (0.3, 0.5]$
	4	$x \in (0.5, 1.0]$
Share of Votes	0	$x = 0$
	1	$x \in (0.0, 0.1]$
	2	$x \in (0.1, 0.25]$
	3	$x \in (0.25, 0.4]$
	4	$x \in (0.4, 1.0]$
Share of Parliamentary Seats	0	$x = 0$
	1	$x \in (0.0, 0.1]$
	2	$x \in (0.1, 0.25]$
	3	$x \in (0.25, 0.4]$
	4	$x \in (0.4, 1.0]$
Ratio Votes	0	$x = 0$
	1	$x \in (0.0, 0.2]$
	2	$x \in (0.2, 0.4]$
	3	$x \in (0.4, 0.6]$
	4	$x \in (0.6, 1.0]$
Ratio Parliamentary Seats	0	$x = 0$
	1	$x \in (0.0, 0.2]$
	2	$x \in (0.2, 0.4]$
	3	$x \in (0.4, 0.6]$
	4	$x \in (0.6, 1.0]$

Notes: The item-response approach requires ordinal regime characteristics. This table indicates how we transform our continuous regime characteristics into ordinal regime characteristics.

Table C.2 Consequences of using different aggregation functions — 2SLS estimates (varying instruments).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second-stage estimates					
Democracy	0.030^{***}	0.055^{***}	0.073^{***}	0.039^{***}	0.045^{***}	0.037^{***}
	(0.006)	(0.010)	(0.013)	(0.007)	(0.008)	(0.007)
Equal. (p-val.)	–	0.000	0.000	0.146	0.013	0.238
	Panel B: First-stage estimates					
Demo. (reg.)	0.740^{***}	0.607^{***}	0.515^{***}	0.757^{***}	0.709^{***}	0.755^{***}
	(0.064)	(0.069)	(0.061)	(0.059)	(0.064)	(0.059)
Equal. (p-val.)	–	0.039	0.001	0.790	0.620	0.826
Observations	10026	10026	10026	10026	10026	10026
Countries	163	163	163	163	163	163
SW (F-stat.)	133.77	77.01	70.48	164.79	121.10	161.88
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.708	2.828	3.473	2.119	2.387	2.060

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) concerns the instrumental variable: while the baseline analysis uses the same instrument in all specifications, this robustness check allows for changes in the instrumental variable due to changes in the aggregation method.

Table C.3 Consequences of using different aggregation functions — OLS estimates (5-year data).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.078^{***}	0.150^{***}	0.170^{***}	0.104^{***}	0.124^{***}	0.101^{***}
	(0.017)	(0.029)	(0.031)	(0.022)	(0.025)	(0.022)
Income_{t-1}	0.933^{***}	0.927^{***}	0.927^{***}	0.931^{***}	0.929^{***}	0.932^{***}
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	2116	2116	2116	2116	2116	2116
Countries	163	163	163	163	163	163
R-Squared	0.915	0.915	0.916	0.915	0.915	0.915
F Stat	7058	7116	7074	6970	7050	6929
Equal. (p-val.)	–	0.000	0.000	0.126	0.008	0.178
Long-run effect	1.164	2.063	2.321	1.513	1.759	1.477

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 2) is that we use five-year averages rather than annual data.

Table C.4 Consequences of using different aggregation functions — 2SLS estimates, (5-year data).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Second-stage estimates			
Democracy	0.117^{***}	0.196^{***}	0.236^{***}	0.141^{***}	0.160^{***}	0.138^{***}
	(0.032)	(0.052)	(0.062)	(0.038)	(0.043)	(0.038)
Income_{t-1}	0.925^{***}	0.921^{***}	0.918^{***}	0.925^{***}	0.923^{***}	0.925^{***}
	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.010)
Equal. (p-val.)	–	0.013	0.000	0.456	0.170	0.514
			Panel B: First-stage estimates			
Demo. (reg.)	0.989^{***}	0.580^{***}	0.823^{***}	0.785^{***}	0.721^{***}	0.840^{***}
	(0.088)	(0.052)	(0.072)	(0.062)	(0.062)	(0.073)
Equal. (p-val.)	–	0.000	0.000	0.059	0.003	0.091
Observations	2116	2116	2116	2116	2116	2116
Countries	163	163	163	163	163	163
R-Squared	0.915	0.915	0.915	0.915	0.915	0.915
F Stat	6456	6455	6369	6374	6413	6352
SW (F-stat.)	127.71	130.06	129.30	129.63	132.35	129.74
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.562	2.472	2.882	1.872	2.094	1.841

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) is that we use five-year averages rather than annual data.

Table C.5 Consequences of using different aggregation functions — OLS estimates (with controls).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.015^{***}	0.032^{***}	0.029^{***}	0.020^{***}	0.025^{***}	0.020^{***}
	(0.003)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)
Observations	9654	9654	9654	9654	9654	9654
Countries	161	161	161	161	161	161
R-Squared	0.987	0.987	0.987	0.987	0.987	0.987
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	0.000	0.000	0.167	0.006	0.210
Long-run effect	1.058	2.061	1.901	1.364	1.658	1.338

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 2) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.6 Consequences of using different aggregation functions — 2SLS estimates (with controls).

	Machine Learning	**Additive**	**Item-Response**	**Multi-plicative**	**Add./ Multi. (Average)**	**Add./ Multi. (CD function)**
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Second-stage estimates			
Democracy	0.033^{***}	0.057^{***}	0.067^{***}	0.039^{***}	0.045^{***}	0.038^{***}
	(0.007)	(0.012)	(0.015)	(0.009)	(0.010)	(0.008)
Equal. (p-val.)	–	0.000	0.000	0.428	0.092	0.505
			Panel B: First-stage estimates			
Demo. (reg.)	0.752^{***}	0.428^{***}	0.367^{***}	0.639^{***}	0.547^{***}	0.655^{***}
	(0.077)	(0.045)	(0.039)	(0.065)	(0.055)	(0.066)
Equal. (p-val.)	–	0.000	0.000	0.149	0.009	0.213
Observations	9654	9654	9654	9654	9654	9654
Countries	161	161	161	161	161	161
Controls	Yes	Yes	Yes	Yes	Yes	Yes
SW (F-stat.)	94.36	92.56	88.96	97.62	98.00	97.97
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.993	3.264	3.624	2.318	2.654	2.276

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.7 Consequences of using different aggregation functions — OLS estimates (property rights).

	Machine Learning	**Additive**	**Item-Response**	**Multi-plicative**	**Add./ Multi. (Average)**	**Add./ Multi. (CD function)**
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.062^{***}	0.133^{***}	0.074^{***}	0.070^{***}	0.094^{***}	0.068^{***}
	(0.007)	(0.013)	(0.009)	(0.008)	(0.010)	(0.008)
Observations	11565	11565	11565	11565	11565	11565
Countries	175	175	175	175	175	175
R-Squared	0.912	0.916	0.910	0.911	0.913	0.911
Equal. (p-val.)	–	0.000	0.097	0.291	0.000	0.393
Long-run effect	0.447	0.760	0.664	0.512	0.617	0.506

Notes: This table presents OLS estimates. The dependent variable is an expert-based measure of private property protection. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 2) is the dependent variable.

Table C.8 Consequences of using different aggregation functions — 2SLS estimates (property rights).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second-stage estimates					
Democracy	0.081^{***}	0.144^{***}	0.145^{***}	0.098^{***}	0.114^{***}	0.097^{***}
	(0.014)	(0.025)	(0.026)	(0.016)	(0.019)	(0.016)
Equal. (p-val.)	–	0.000	0.000	0.428	0.092	0.505
	Panel B: First-stage estimates					
Demo. (reg.)	0.499^{***}	0.279^{***}	0.278^{***}	0.410^{***}	0.353^{***}	0.416^{***}
	(0.061)	(0.031)	(0.031)	(0.047)	(0.040)	(0.048)
Equal. (p-val.)	–	0.001	0.001	0.201	0.028	0.229
Observations	11565	11565	11565	11565	11565	11565
Countries	175	175	175	175	175	175
SW (F-stat.)	66.86	80.34	81.60	75.47	79.41	74.29
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	0.500	0.781	0.904	0.593	0.663	0.584

Notes: This table presents 2SLS estimates. The dependent variable is an expert-based measure of private property protection. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) is the dependent variable.

Table C.9 Consequences of using different aggregation functions — OLS estimates (education).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.208^{***}	0.372^{***}	0.353^{***}	0.253^{***}	0.305^{***}	0.248^{***}
	(0.051)	(0.079)	(0.077)	(0.061)	(0.069)	(0.059)
Observations	2107	2107	2107	2107	2107	2107
Countries	147	147	147	147	147	147
R-Squared	0.977	0.977	0.977	0.977	0.977	0.977
Equal. (p-val.)	–	0.002	0.005	0.379	0.060	0.431

Notes: This table presents OLS estimates. The dependent variable is the average years of schooling. Since annual data of the dependent variable does not exist, we use five-year data. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The main difference compared to our baseline analysis (see Table 2) is the dependent variable.

Table C.10 Consequences of using different aggregation functions — 2SLS estimates (education).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Second-stage estimates			
Democracy	0.616***	1.013***	1.259***	0.752***	0.847***	0.732***
	(0.135)	(0.211)	(0.275)	(0.169)	(0.185)	(0.165)
Equal. (p-val.)	–	0.003	0.000	0.316	0.088	0.392
			Panel B: First-stage estimates			
Demo. (reg.)	0.867***	0.527***	0.424***	0.710***	0.630***	0.730***
	(0.013)	(0.060)	(0.052)	(0.081)	(0.071)	(0.083)
Equal. (p-val.)	–	0.001	0.000	0.104	0.015	0.154
Observations	2107	2107	2107	2107	2107	2107
Countries	147	147	147	147	147	147
SW (F-stat.)	82.12	77.59	67.23	77.05	79.78	78.20
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	44.47	54.81	53.46	50.37	52.22	51.66

Notes: This table presents 2SLS estimates. The dependent variable is the average years of schooling. Since annual data of the dependent variable does not exist, we use five-year data. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The main difference compared to our baseline analysis (see Table 3) is the dependent variable.

Table C.11 Consequences of using different aggregation functions — OLS estimates (Alternative Concept I).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.014***	0.020***	0.020***	0.019***	0.020***	0.019***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Observations	10,026	10,026	10,026	10,026	10,026	10,026
Countries	163	163	163	163	163	163
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985
GDP Dynamics	Yes	Yes	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	0.022	0.033	0.052	0.032	0.055
Long-run effect	0.987	1.396	1.351	1.326	1.364	1.320

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 2) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of only one aspect in this robustness check (political competition).

Table C.12 Consequences of using different aggregation functions — 2SLS estimates (Alternative Concept I).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Second-stage estimates			
Democracy	0.032***	0.046***	0.045***	0.044***	0.045***	0.043***
	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Equal. (p-val.)	–	0.019	0.026	0.041	0.029	0.055
			Panel B: First-stage estimates			
Demo. (reg.)	0.907***	0.628***	0.638***	0.654***	0.641***	0.664***
	(0.081)	(0.055)	(0.056)	(0.058)	(0.056)	(0.059)
Equal. (p-val.)	–	0.001	0.001	0.002	0.001	0.003
Observations	10026	10026	10026	10026	10026	10026
Countries	163	163	163	163	163	163
SW (F-stat.)	126.22	130.56	129.42	128.46	129.72	128.98
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.828	2.546	2.457	2.424	2.489	2.395

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of only one aspect in this robustness check (political competition).

Table C.13 Consequences of using different aggregation functions — OLS estimates (Alternative Concept II).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.019***	0.033***	0.040***	0.022***	0.027***	0.021***
	(0.003)	(0.006)	(0.007)	(0.004)	(0.005)	(0.004)
Observations	9949	9949	9949	9949	9949	9949
Countries	161	161	161	161	161	161
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985
F Stat	83743	85176	85988	83530	84698	82884
GDP Dynamics	Yes	Yes	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	0.000	0.000	0.251	0.012	0.388
Long-run effect	1.206	2.056	2.511	1.451	1.708	1.402

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 2) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of four aspects in this robustness check (political competition, political participation, freedom of opinion, judiciary independence).

Table C.14 Consequences of using different aggregation functions — 2SLS estimates (Alternative Concept II).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Second-stage estimates			
Democracy	0.036***	0.061***	0.083***	0.043***	0.049***	0.041***
	(0.007)	(0.012)	(0.015)	(0.008)	(0.009)	(0.008)
Equal. (p-val.)	–	0.000	0.000	0.328	0.061	0.414
			Panel B: First-stage estimates			
Demo. (reg.)	0.901***	0.531***	0.388***	0.757***	0.661***	0.777***
	(0.095)	(0.056)	(0.039)	(0.079)	(0.069)	(0.081)
Equal. (p-val.)	–	0.000	0.000	0.132	0.012	0.196
Observations	9949	9949	9949	9949	9949	9949
Countries	161	161	161	161	161	161
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985
F Stat (Sec.)	70044	70325	65073	70489	70786	70531
SW (F-stat.)	90.20	88.63	100.74	108.07	92.73	92.47
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.915	3.115	4.067	2.284	2.577	2.243

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of four aspects in this robustness check (political competition, political participation, freedom of opinion, judiciary independence).

Table C.15 Consequences of using different aggregation functions — OLS estimates (alternative regime characteristics).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.018***	0.026***	0.033***	0.030***	0.031***	0.026***
	(0.003)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Observations	9935	9935	9935	9935	9935	9935
Countries	161	161	161	161	161	161
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985
GDP Dynamics	Yes	Yes	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	0.018	0.000	0.001	0.000	0.033
Long-run effect	1.778	1.691	2.057	1.692	1.822	1.508

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 2) is that we use the regime characteristics proposed by Teorell et al. (2019) in this robustness check.

Table C.16 Consequences of using different aggregation functions — 2SLS estimates (alternative regime characteristics).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second-stage estimates					
Democracy	0.035***	0.051***	0.052***	0.056***	0.053***	0.045***
	(0.006)	(0.009)	(0.009)	(0.010)	(0.009)	(0.008)
Equal. (p-val.)	–	0.013	0.007	0.001	0.004	0.114
	Panel B: First-stage estimates					
Demo. (reg.)	0.963***	0.666***	0.650***	0.605***	0.636***	0.750***
	(0.094)	(0.065)	(0.065)	(0.063)	(0.061)	(0.073)
Equal. (p-val.)	–	0.002	0.001	0.000	0.001	0.025
Observations	9935	9935	9935	9935	9935	9935
Countries	161	161	161	161	161	161
SW (F-stat.)	104.59	104.86	100.56	91.92	107.80	105.36
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	1.863	2.643	2.650	2.444	2.540	2.149

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. All democracy indices are continuous and range from 0 to 1. The only difference between the six columns is that we used different aggregation method for the creation of the democracy indices. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 6 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 3) is that we use the regime characteristics proposed by Teorell et al. (2019) in this robustness check.

Table C.17 Consequences of using different numerical forms — Machine learning indices (5-year data).

	Continuous SVM index		Dichotomous SVM index	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Democracy	0.078***	0.117***	0.050***	0.087***
	(0.017)	(0.032)	(0.012)	(0.026)
Observations	2116	2116	2116	2116
Countries	163	163	163	163
R-Squared	0.915	0.915	0.914	0.914
F Stat (sec.)	7058	6456	6905	6407
GDP Dynamics	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.108	0.349
First-stage	–	0.989***	–	0.757***
SW (F-stat.)	–	127.71	–	122.67
AR (p-val.)	–	0.000	–	0.000
Long-run effect	1.165	1.562	0.816	1.257

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) is that we use five-year averages rather than annual data.

Table C.18 Consequences of using different numerical forms — Machine learning indices (with controls).

	Continuous SVM index		Dichotomous SVM index	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Democracy	0.015^{***} (0.003)	0.033^{***} (0.007)	0.008^{***} (0.002)	0.029^{***} (0.007)
Observations	9654	9654	9654	9654
Countries	161	161	161	161
R-Squared	0.987	0.987	0.987	0.986
F Stat (sec.)	45586	40906	46048	37668
GDP Dynamics	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.051	0.649
First-stage	–	0.752^{***}	–	0.556^{***}
SW (F-stat.)	–	94.36	–	80.61
AR (p-val.)	–	0.000	–	0.000
Long-run effect	1.058	1.993	0.619	1.859

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.19 Consequences of using different numerical forms — Machine learning indices (property rights).

	Continuous SVM index		Dichotomous SVM index	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Property	0.062^{***} (0.007)	0.081^{***} (0.014)	0.033^{***} (0.004)	0.070^{***} (0.014)
Observations	11565	11565	11565	11565
Countries	175	175	175	175
R-Squared	0.912	0.911	0.910	0.905
F Stat (sec.)	3363	2925	3589	2907
GDP Dynamics	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.000	0.419
First-stage	–	0.449^{***}	–	0.396^{***}
SW (F-stat.)	–	66.86	–	47.11
AR (p-val.)	–	0.000	–	0.000
Long-run effect	0.447	0.498	0.314	0.446

Notes: This table presents OLS and 2SLS estimates. The dependent variable is an expert-based measure of private property protection. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) is the dependent variable.

Table C.20 Consequences of using different numerical forms — Machine learning indices (education).

	Continuous SVM index		Dichotomous SVM index	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Property	0.208^{***} (0.051)	0.616^{***} (0.135)	0.153^{***} (0.037)	0.411^{***} (0.102)
Observations	2107	2107	2107	2107
Countries	147	147	147	147
R-Squared	0.977	0.976	0.977	0.976
F Stat (sec.)	23676	20061	23534	21307
GDP Dynamics	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.283	0.130
First-stage	–	0.868^{***}	–	0.652^{***}
SW (F-stat.)	–	82.12	–	79.23
AR (p-val.)	–	0.000	–	0.000

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the average years of schooling. Since annual data of the dependent variable does not exist, we use five-year data. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) is the dependent variable.

Table C.21 Consequences of using different numerical forms — Machine learning indices (alternative concept I).

	Continuous SVM index		Dichotomous SVM index	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Democracy	0.014^{***} (0.003)	0.033^{***} (0.006)	0.009^{***} (0.002)	0.025^{***} (0.006)
Observations	10026	10026	10026	10026
Countries	163	163	163	163
R-Squared	0.985	0.985	0.985	0.985
F Stat (sec.)	77460	68940	74776	70337
GDP Dynamics	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.046	0.281
First-stage	–	0.907^{***}	–	0.659^{***}
SW (F-stat.)	–	126.22	–	111.06
AR (p-val.)	–	0.000	–	0.000
Long-run effect	0.987	1.828	0.650	1.552

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of only one aspects in this robustness check (political competition).

Table C.22 Consequences of using different numerical forms — Machine learning indices (alternative concept II).

	Continuous SVM index		**Dichotomous SVM index**	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Democracy	0.019^{***}	0.036^{***}	0.014^{***}	0.030^{***}
	(0.003)	(0.007)	(0.003)	(0.006)
Observations	9949	9949	9949	9949
Countries	161	161	161	161
R-Squared	0.985	0.985	0.985	0.985
F Stat (sec.)	78591	68988	74300	68931
GDP Dynamics	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.161	0.418
First-stage	–	0.901^{***}	–	0.658^{***}
SW (F-stat.)	–	90.20	–	81.04
AR (p-val.)	–	0.000	–	0.000
Long-run effect	1.206	1.915	0.939	1.659

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of four aspects in this robustness check (political competition, political participation, freedom of opinion, judiciary independence).

Table C.23 Consequences of using different numerical forms — Machine learning indices (alternative regime characteristics).

	Continuous SVM index		**Dichotomous SVM index**	
	(1) *OLS*	(2) *2SLS*	(3) *OLS*	(4) *2SLS*
Democracy	0.018^{***}	0.034^{***}	0.012^{***}	0.026^{***}
	(0.003)	(0.006)	(0.002)	(0.005)
Observations	9935	9935	9935	9935
Countries	161	161	161	161
R-Squared	0.985	0.985	0.985	0.985
F Stat (sec.)	86947	70590	85823	70071
GDP Dynamics	Yes	Yes	Yes	Yes
Equal. (p-val.)	–	–	0.095	0.218
First-stage	–	0.881^{***}	–	0.732^{***}
SW (F-stat.)	–	126.12	–	113.95
AR (p-val.)	–	0.000	–	0.000
Long-run effect	1.178	1.825	0.842	1.472

Notes: This table presents OLS and 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between Column 1/2 and Column 3/4 is the numerical form of the machine learning index. In the 2SLS regressions, the regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Column 3/4 are significantly different from the estimates reported in Column 1/2. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Table 5) is that we use the regime characteristics proposed by Teorell et al. (2019) in this robustness check.

Table C.24 Consequences of using different numerical forms — Threshold approach, multiplicative index, OLS estimates.

	Continuous (Multi.)	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Democracy	0.23^{***}	0.009^{***}	0.009^{***}	0.013^{***}	0.014^{***}	0.014^{***}	0.012^{***}
	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Observations	10,026	10,026	10,026	10,026	10,026	10,026	10,026
Countries	163	163	163	163	163	163	163
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985	0.985
Equal. (p-val.)	–	0.001	0.001	0.014	0.032	0.023	0.006
Long-run effect	1.446	0.392	0.687	0.894	0.936	0.911	0.860

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

Table C.25 Consequences of using different numerical forms — Threshold approach, multiplicative index, 2SLS estimates.

	Continuous (Multi.)	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Second-stage estimates						
Democracy	0.039^{***}	0.026^{***}	0.027^{***}	0.029^{***}	0.034^{***}	0.056^{***}	0.135^{***}
	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)	(0.012)	(0.041)
Equal. (p-val.)	–	0.063	0.093	0.164	0.496	0.016	0.000
	Panel B: First-stage estimates						
Demo. (reg.)	0.757^{***}	1.145^{***}	1.092^{***}	1.015^{***}	0.864^{***}	0.526^{***}	0.219^{***}
	(0.059)	(0.090)	(0.096)	(0.085)	(0.084)	(0.078)	(0.056)
Equal. (p-val.)	–	0.000	0.000	0.000	0.070	0.000	0.000
Observations	10026	10026	10026	10026	10026	10026	10026
Countries	163	163	163	163	163	163	163
R-Squared	0.985	0.985	0.985	0.985	0.985	0.985	0.985
GDP Dynamics	yes	yes	yes	yes	yes	yes	yes
F Stat (Sec.)	69257	68285	68638	64219	64902	49404	16944
SW (F-stat.)	164.79	160.24	130.40	141.02	106.64	45.64	15.40
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Long-run effect	2.119	1.596	1.582	1.593	1.684	2.184	3.729

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table 3 because we only use the multiplicative index to compute the regional degree of democratization. Our results do not change if we use the original instruments.

Table C.26 Consequences of using different numerical forms — Threshold approach, several indices, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.032^{***}	0.010^{***}	0.009^{***}	0.011^{***}	0.012^{***}	0.013^{***}	0.012^{***}
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Equal. (p-val.)	–	0.000	0.000	0.000	0.000	0.001	0.000
	Panel B: Item-Response Approach						
Democracy	0.033^{***}	0.014^{***}	0.012^{***}	0.010^{***}	0.012^{***}	0.007^{***}	0.012^{***}
	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.001	0.000	0.000	0.000	0.000	0.000
	Panel C: Additive/ Multiplicative Approach (Weighted Average)						
Democracy	0.027^{***}	0.009^{***}	0.009^{***}	0.011^{***}	0.013^{***}	0.014^{***}	0.012^{***}
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Equal. (p-val.)	–	0.000	0.000	0.001	0.002	0.005	0.001
	Panel D: Additive/ Multiplicative Approach (CD function)						
Democracy	0.022^{***}	0.009^{***}	0.010^{***}	0.011^{***}	0.014^{***}	0.014^{***}	0.012^{***}
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Equal. (p-val.)	–	0.001	0.001	0.004	0.029	0.044	0.007

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.27 Consequences of using different numerical forms — Threshold approach, several indices, 2SLS estimates.

	Continuous	**Threshold (0.3)**	**Threshold (0.4)**	**Threshold (0.5)**	**Threshold (0.6)**	**Threshold (0.7)**	**Threshold (0.8)**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: Additive Approach			
Democracy	0.055***	0.041***	0.028***	0.027***	0.031***	0.047***	0.081***
	(0.010)	(0.009)	(0.006)	(0.005)	(0.006)	(0.010)	(0.019)
Equal. (p-val.)	–	0.154	0.008	0.006	0.017	0.41	0.012
				Panel B: Item-Response Approach			
Democracy	0.073***	0.082***	0.046***	0.031***	0.037***	0.057***	0.108***
	(0.013)	(0.017)	(0.009)	(0.006)	(0.007)	(0.012)	(0.026)
Equal. (p-val.)	–	0.478	0.034	0.001	0.005	0.220	0.006
				Panel C: Additive/ Multiplicative Approach (Weighted Average)			
Democracy	0.045***	0.026***	0.027***	0.028***	0.032***	0.052***	0.104***
	(0.008)	(0.005)	(0.005)	(0.005)	(0.006)	(0.011)	(0.028)
Equal. (p-val.)	–	0.024	0.025	0.040	0.122	0.367	0.000
				Panel D: Additive/ Multiplicative Approach (CD function)			
Democracy	0.037***	0.026***	0.026***	0.028***	0.032***	0.051***	0.103***
	(0.007)	(0.005)	(0.005)	(0.005)	(0.006)	(0.011)	(0.028)
Equal. (p-val.)	–	0.090	0.106	0.173	0.454	0.047	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table 3 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments.

Table C.28 Consequences of using different numerical forms — Threshold approach, 5-year data, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.150^{***}	0.033^{***}	0.037^{***}	0.043^{***}	0.055^{***}	0.070^{***}	0.064^{***}
	(0.029)	(0.013)	(0.011)	(0.011)	(0.012)	(0.015)	(0.014)
Equal. (p-val.)	–	0.000	0.000	0.000	0.001	0.006	0.003
	Panel B: Item-Response Approach						
Democracy	0.170^{***}	0.048^{***}	0.046^{***}	0.041^{***}	0.049^{***}	0.063^{***}	0.066^{***}
	(0.031)	(0.020)	(0.013)	(0.012)	(0.011)	(0.012)	(0.015)
Equal. (p-val.)	–	0.000	0.000	0.000	0.000	0.001	0.001
	Panel C: Multiplicative Approach						
Democracy	0.104^{***}	0.040^{***}	0.054^{***}	0.059^{***}	0.079^{***}	0.071^{***}	0.063^{***}
	(0.022)	(0.011)	(0.011)	(0.011)	(0.013)	(0.014)	(0.015)
Equal. (p-val.)	–	0.003	0.021	0.037	0.252	0.130	0.061
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.124^{***}	0.032^{***}	0.041^{***}	0.055^{***}	0.071^{***}	0.076^{***}	0.064^{***}
	(0.025)	(0.012)	(0.011)	(0.012)	(0.013)	(0.015)	(0.015)
Equal. (p-val.)	–	0.000	0.001	0.006	0.032	0.054	0.016
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.101^{***}	0.037^{***}	0.045^{***}	0.057^{***}	0.071^{***}	0.078^{***}	0.064^{***}
	(0.022)	(0.011)	(0.011)	(0.012)	(0.013)	(0.015)	(0.015)
Equal. (p-val.)	–	0.004	0.011	0.041	0.161	0.280	0.089

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26) is that we use five-year averages rather than annual data.

Table C.29 Consequences of using different numerical forms — Threshold approach, 5-year data, 2LS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.199***	0.146***	0.097***	0.093***	0.112***	0.167***	0.327***
	(0.052)	(0.042)	(0.027)	(0.025)	(0.029)	(0.046)	(0.101)
Equal. (p-val.)	–	0.305	0.050	0.040	0.093	0.546	0.013
	Panel B: Item-Response Approach						
Democracy	0.256***	0.299***	0.152***	0.104***	0.123***	0.202***	0.400***
	(0.063)	(0.081)	(0.040)	(0.027)	(0.032)	(0.053)	(0.117)
Equal. (p-val.)	–	0.501	0.097	0.016	0.035	0.391	0.022
	Panel C: Multiplicative Approach						
Democracy	0.140***	0.090***	0.093***	0.102***	0.123***	0.199***	0.605***
	(0.039)	(0.025)	(0.025)	(0.028)	(0.033)	(0.059)	(0.252)
Equal. (p-val.)	–	0.192	0.222	0.317	0.662	0.128	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.162***	0.093***	0.095***	0.100***	0.120***	0.188***	0.493***
	(0.043)	(0.026)	(0.026)	(0.032)	(0.032)	(0.054)	(0.181)
Equal. (p-val.)	–	0.107	0.117	0.148	0.333	0.549	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.135***	0.089***	0.091***	0.099***	0.117***	0.187***	0.480***
	(0.038)	(0.026)	(0.026)	(0.028)	(0.032)	(0.057)	(0.183)
Equal. (p-val.)	–	0.227	0.254	0.347	0.639	0.173	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.4 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) is that we use five-year averages rather than annual data.

Table C.30 Consequences of using different numerical forms — Threshold approach, with controls, OLS estimates.

	Continuous	**Threshold (0.3)**	**Threshold (0.4)**	**Threshold (0.5)**	**Threshold (0.6)**	**Threshold (0.7)**	**Threshold (0.8)**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.032^{***}	0.009^{***}	0.007^{***}	0.011^{***}	0.010^{***}	0.008^{***}	0.006^{***}
	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Equal. (p-val.)	–	0.001	0.000	0.001	0.001	0.000	0.000
	Panel B: Item-Response Approach						
Democracy	0.029^{***}	0.013^{***}	0.011^{***}	0.009^{***}	0.008^{***}	0.004^{***}	0.006^{***}
	(0.006)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.011	0.006	0.002	0.001	0.000	0.000
	Panel C: Multiplicative Approach						
Democracy	0.020^{***}	0.007^{***}	0.008^{***}	0.011^{***}	0.010^{***}	0.009^{***}	0.006^{***}
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Equal. (p-val.)	–	0.003	0.005	0.042	0.028	0.010	0.002
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.025^{***}	0.008^{***}	0.008^{***}	0.010^{***}	0.010^{***}	0.010^{***}	0.006^{***}
	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Equal. (p-val.)	–	0.001	0.001	0.005	0.004	0.005	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.020^{***}	0.008^{***}	0.008^{***}	0.010^{***}	0.010^{***}	0.010^{***}	0.006^{***}
	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Equal. (p-val.)	–	0.006	0.006	0.031	0.024	0.031	0.002

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.31 Consequences of using different numerical forms — Threshold approach, with controls, 2SLS estimates.

	Continuous	**Threshold (0.3)**	**Threshold (0.4)**	**Threshold (0.5)**	**Threshold (0.6)**	**Threshold (0.7)**	**Threshold (0.8)**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.057^{***}	0.041^{***}	0.028^{***}	0.027^{***}	0.032^{***}	0.056^{***}	0.091^{***}
	(0.012)	(0.010)	(0.007)	(0.006)	(0.007)	(0.014)	(0.026)
Equal. (p-val.)	–	0.208	0.020	0.017	0.050	0.974	0.006
	Panel B: Item-Response Approach						
Democracy	0.069^{***}	0.072^{***}	0.039^{***}	0.027^{***}	0.036^{***}	0.061^{***}	0.126^{***}
	(0.014)	(0.020)	(0.009)	(0.006)	(0.009)	(0.014)	(0.037)
Equal. (p-val.)	–	0.792	0.038	0.004	0.023	0.572	0.000
	Panel C: Multiplicative Approach						
Democracy	0.039^{***}	0.024^{***}	0.027^{***}	0.030^{***}	0.037^{***}	0.075^{***}	0.168^{***}
	(0.009)	(0.006)	(0.006)	(0.007)	(0.008)	(0.019)	(0.068)
Equal. (p-val.)	–	0.089	0.166	0.281	0.847	0.000	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.046^{***}	0.025^{***}	0.026^{***}	0.029^{***}	0.034^{***}	0.067^{***}	0.124^{***}
	(0.010)	(0.006)	(0.006)	(0.007)	(0.007)	(0.017)	(0.042)
Equal. (p-val.)	–	0.038	0.052	0.090	0.265	0.028	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.038^{***}	0.024^{***}	0.026^{***}	0.029^{***}	0.035^{***}	0.067^{***}	0.125^{***}
	(0.008)	(0.006)	(0.006)	(0.007)	(0.008)	(0.017)	(0.043)
Equal. (p-val.)	–	0.116	0.174	0.295	0.746	0.001	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.6 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.32 Consequences of using different numerical forms — Threshold approach, roperty rights, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.133^{***}	0.033^{***}	0.037^{***}	0.035^{***}	0.038^{***}	0.032^{***}	0.016^{***}
	(0.013)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Equal. (p-val.)	–	0.000	0.000	0.000	0.000	0.000	0.000
	Panel B: Item-Response Approach						
Democracy	0.074^{***}	0.023^{***}	0.022^{***}	0.025^{***}	0.024^{***}	0.015^{***}	0.013^{***}
	(0.009)	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)
Equal. (p-val.)	–	0.000	0.000	0.000	0.000	0.000	0.000
	Panel C: Multiplicative Approach						
Democracy	0.070^{***}	0.030^{***}	0.032^{***}	0.034^{***}	0.032^{***}	0.026^{***}	0.011^{***}
	(0.008)	(0.040)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
Equal. (p-val.)	–	0.000	0.000	0.000	0.000	0.000	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.094^{***}	0.030^{***}	0.032^{***}	0.034^{***}	0.035^{***}	0.030^{***}	0.012^{***}
	(0.010)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
Equal. (p-val.)	–	0.000	0.001	0.002	0.000	0.000	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.068^{***}	0.029^{***}	0.032^{***}	0.034^{***}	0.035^{***}	0.030^{***}	0.012^{***}
	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)
Equal. (p-val.)	–	0.000	0.001	0.002	0.000	0.000	0.000

Notes: This table presents OLS estimates. The dependent variable is an expert-based measure of private property protection. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26) is that we change the outcome variable.

Table C.33 Consequences of using different numerical forms — Threshold approach, property rights, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.145***	0.131***	0.085***	0.072***	0.080***	0.121***	0.129***
	(0.025)	(0.036)	(0.019)	(0.015)	(0.014)	(0.023)	(0.030)
Equal. (p-val.)	–	0.586	0.020	0.004	0.011	0.355	0.533
	Panel B: Item-Response Approach						
Democracy	0.118***	0.134***	0.078***	0.064***	0.073***	0.096***	0.141***
	(0.023)	(0.052)	(0.019)	(0.014)	(0.017)	(0.022)	(0.038)
Equal. (p-val.)	–	0.472	0.081	0.018	0.052	0.351	0.297
	Panel C: Multiplicative Approach						
Democracy	0.078***	0.056***	0.057***	0.059***	0.063***	0.100***	0.134***
	(0.012)	(0.010)	(0.010)	(0.010)	(0.011)	(0.023)	(0.039)
Equal. (p-val.)	–	0.079	0.070	0.109	0.276	0.006	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.115***	0.075***	0.070***	0.072***	0.077***	0.131***	0.159***
	(0.019)	(0.015)	(0.014)	(0.014)	(0.013)	(0.024)	(0.039)
Equal. (p-val.)	–	0.037	0.020	0.026	0.048	0.385	0.021
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.094***	0.070***	0.067***	0.071***	0.077***	0.134***	0.164***
	(0.015)	(0.013)	(0.013)	(0.014)	(0.013)	(0.025)	(0.040)
Equal. (p-val.)	–	0.112	0.082	0.132	0.265	0.009	0.000

Notes: This table presents 2SLS estimates. The dependent variable is an expert-based measure of private property protection. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.8 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) is that we change the outcome variable.

Table C.34 Consequences of using different numerical forms — Threshold approach, education, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.372^{***}	0.123^{***}	0.156^{***}	0.151^{***}	0.130^{***}	0.049	0.032
	(0.079)	(0.029)	(0.031)	(0.032)	(0.037)	(0.037)	(0.037)
Equal. (p-val.)	–	0.002	0.007	0.006	0.003	0.000	0.000
	Panel B: Item-Response Approach						
Democracy	0.353^{***}	0.135^{***}	0.121^{***}	0.141^{***}	0.124^{***}	0.041	-0.016
	(0.077)	(0.034)	(0.030)	(0.029)	(0.030)	(0.033)	(0.038)
Equal. (p-val.)	–	0.006	0.003	0.007	0.004	0.000	0.000
	Panel C: Multiplicative Approach						
Democracy	0.253^{***}	0.138^{***}	0.155^{***}	0.133^{***}	0.078^{***}	0.025	0.036
	(0.061)	(0.030)	(0.030)	(0.033)	(0.038)	(0.039)	(0.044)
Equal. (p-val.)	–	0.059	0.108	0.050	0.005	0.000	0.001
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.305^{***}	0.129^{***}	0.171^{***}	0.134^{***}	0.113^{***}	0.042	0.057
	(0.069)	(0.029)	(0.031)	(0.033)	(0.037)	(0.038)	(0.038)
Equal. (p-val.)	–	0.012	0.054	0.014	0.006	0.000	0.001
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.248^{***}	0.133^{***}	0.164^{***}	0.126^{***}	0.108^{***}	0.042	0.057
	(0.059)	(0.029)	(0.031)	(0.031)	(0.037)	(0.037)	(0.038)
Equal. (p-val.)	–	0.052	0.155	0.041	0.019	0.001	0.002

Notes: This table presents OLS estimates. The dependent variable is the average years of schooling. Since annual data of the dependent variable does not exist, we use five-year data. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26) is that we change the outcome variable.

Table C.35 Consequences of using different numerical forms — Threshold approach, education, 2SLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	1.273^{***}	0.874^{***}	0.590^{***}	0.610^{***}	0.743^{***}	1.110^{***}	3.804
	(0.240)	(0.191)	(0.108)	(0.112)	(0.149)	(0.288)	(2.524)
Equal. (p-val.)	–	0.097	0.005	0.006	0.027	0.498	0.000
	Panel B: Item-Response Approach						
Democracy	1.780^{***}	1.616^{***}	0.997^{***}	0.725^{***}	0.897^{***}	1.826^{***}	8.408
	(0.325)	(0.378)	(0.206)	(0.128)	(0.177)	(0.177)	(9.163)
Equal. (p-val.)	–	0.614	0.016	0.001	0.007	0.886	0.000
	Panel C: Multiplicative Approach						
Democracy	0.640^{***}	0.378^{***}	0.403^{***}	0.452^{***}	0.551^{***}	0.951^{***}	4.080
	(0.156)	(0.096)	(0.098)	(0.112)	(0.148)	(0.326)	(3.729)
Equal. (p-val.)	–	0.149	0.200	0.333	0.734	0.026	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.835^{***}	0.454^{***}	0.484^{***}	0.525^{***}	0.631^{***}	1.050^{***}	5.167
	(0.184)	(0.101)	(0.103)	(0.113)	(0.148)	(0.328)	(5.197)
Equal. (p-val.)	–	0.038	0.057	0.092	0.267	0.244	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.590^{***}	0.377^{***}	0.397^{***}	0.446^{***}	0.525^{***}	0.891^{***}	3.903
	(0.154)	(0.098)	(0.096)	(0.111)	(0.138)	(0.302)	(3.627)
Equal. (p-val.)	–	0.166	0.207	0.348	0.670	0.051	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the average years of schooling. Since annual data of the dependent variable does not exist, we use five-year data. All regressions include one lag of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.10 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) is that we change the outcome variable.

Table C.36 Consequences of using different numerical forms — Threshold approach, alternative concept I, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.020^{***}	0.008^{***}	0.011^{***}	0.011^{***}	0.010^{***}	0.007^{***}	0.004^{***}
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.001	0.012	0.013	0.004	0.000	0.000
	Panel B: Item-Response Approach						
Democracy	0.020^{***}	0.013^{***}	0.012^{***}	0.009^{***}	0.011^{***}	0.008^{***}	0.007^{***}
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.037	0.029	0.003	0.017	0.001	0.000
	Panel C: Multiplicative Approach						
Democracy	0.019^{***}	0.009^{***}	0.011^{***}	0.010^{***}	0.008^{***}	0.007^{***}	0.004^{***}
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.004	0.011	0.007	0.001	0.000	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.020^{***}	0.008^{***}	0.011^{***}	0.010^{***}	0.009^{***}	0.006^{***}	0.005^{***}
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.001	0.013	0.008	0.003	0.000	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.019^{***}	0.008^{***}	0.011^{***}	0.010^{***}	0.009^{***}	0.006^{***}	0.005^{***}
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Equal. (p-val.)	–	0.002	0.017	0.010	0.003	0.000	0.000

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of only one aspects in this robustness check (political competition).

Table C.37 Consequences of using different numerical forms — Threshold approach, alternative concept I, 2SLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: Additive Approach			
Democracy	0.046***	0.028***	0.032***	0.034***	0.040***	0.059***	0.174***
	(0.009)	(0.005)	(0.006)	(0.007)	(0.008)	(0.012)	(0.051)
Equal. (p-val.)	–	0.040	0.101	0.162	0.496	0.127	0.000
				Panel B: Item-Response Approach			
Democracy	0.048***	0.042***	0.031***	0.031***	0.034***	0.050***	0.062***
	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.010)	(0.013)
Equal. (p-val.)	–	0.474	0.037	0.043	0.083	0.808	0.093
				Panel C: Multiplicative Approach			
Democracy	0.045***	0.030***	0.032***	0.036***	0.042***	0.066***	0.233***
	(0.008)	(0.006)	(0.006)	(0.007)	(0.008)	(0.014)	(0.077)
Equal. (p-val.)	–	0.077	0.134	0.310	0.730	0.010	0.000
				Panel D: Additive/ Multiplicative Approach (Weigthed Average)			
Democracy	0.045***	0.029***	0.032***	0.035***	0.040***	0.062***	0.208***
	(0.008)	(0.006)	(0.006)	(0.007)	(0.008)	(0.012)	(0.064)
Equal. (p-val.)	–	0.059	0.115	0.223	0.573	0.048	0.000
				Panel E: Additive/ Multiplicative Approach (CD function)			
Democracy	0.044***	0.029***	0.032***	0.035***	0.040***	0.062***	0.209***
	(0.008)	(0.006)	(0.006)	(0.007)	(0.008)	(0.013)	(0.065)
Equal. (p-val.)	–	0.079	0.152	0.284	0.702	0.022	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.12 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of only one aspects in this robustness check (political competition).

Table C.38 Consequences of using different numerical forms — Threshold approach, alternative concept II, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: Additive Approach			
Democracy	0.033***	0.009***	0.008***	0.011***	0.014***	0.013***	0.013***
	(0.006)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Equal. (p-val.)	–	0.000	0.000	0.000	0.002	0.001	0.001
				Panel B: Item-Response Approach			
Democracy	0.040***	0.013***	0.012***	0.011***	0.007***	0.012***	0.006***
	(0.007)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Equal. (p-val.)	–	0.000	0.000	0.000	0.000	0.000	0.000
				Panel C: Multiplicative Approach			
Democracy	0.022***	0.009***	0.009***	0.013***	0.013***	0.012***	0.011***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Equal. (p-val.)	–	0.001	0.001	0.023	0.024	0.009	0.006
				Panel D: Additive/ Multiplicative Approach (Weigthed Average)			
Democracy	0.027***	0.009***	0.009***	0.013***	0.013***	0.013***	0.013***
	(0.005)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Equal. (p-val.)	–	0.000	0.000	0.005	0.005	0.003	0.003
				Panel E: Additive/ Multiplicative Approach (CD function)			
Democracy	0.021***	0.009***	0.009***	0.013***	0.013***	0.013***	0.013***
	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Equal. (p-val.)	–	0.001	0.002	0.040	0.040	0.023	0.026

Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of four aspects in this robustness check (political competition, political participation, freedom of opinion, judiciary independence).

Table C.39 Consequences of using different numerical forms — Threshold approach, alternative concept II, 2SLS estimates.

	Continuous	**Threshold (0.3)**	**Threshold (0.4)**	**Threshold (0.5)**	**Threshold (0.6)**	**Threshold (0.7)**	**Threshold (0.8)**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.059^{***}	0.037^{***}	0.026^{***}	0.029^{***}	0.035^{***}	0.050^{***}	0.114^{***}
	(0.012)	(0.009)	(0.006)	(0.006)	(0.007)	(0.011)	(0.035)
Equal. (p-val.)	–	0.076	0.008	0.013	0.050	0.457	0.000
	Panel B: Item-Response Approach						
Democracy	0.093^{***}	0.079^{***}	0.035^{***}	0.037^{***}	0.059^{***}	0.119^{***}	0.215^{**}
	(0.017)	(0.021)	(0.007)	(0.007)	(0.013)	(0.030)	(0.088)
Equal. (p-val.)	–	0.400	0.001	0.001	0.045	0.118	0.000
	Panel C: Multiplicative Approach						
Democracy	0.042^{***}	0.026^{***}	0.029^{***}	0.035^{***}	0.046^{***}	0.074^{***}	0.188^{***}
	(0.008)	(0.005)	(0.006)	(0.007)	(0.010)	(0.018)	(0.071)
Equal. (p-val.)	–	0.053	0.107	0.381	0.650	0.000	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.048^{***}	0.026^{***}	0.027^{***}	0.032^{***}	0.040^{***}	0.064^{***}	0.158^{***}
	(0.009)	(0.005)	(0.006)	(0.007)	(0.008)	(0.015)	(0.058)
Equal. (p-val.)	–	0.021	0.028	0.098	0.404	0.097	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.040^{***}	0.025^{***}	0.027^{***}	0.032^{***}	0.040^{***}	0.063^{***}	0.159^{***}
	(0.008)	(0.005)	(0.006)	(0.006)	(0.008)	(0.015)	(0.058)
Equal. (p-val.)	–	0.066	0.097	0.310	0.966	0.003	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.14 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) concerns the concept of democracy: while our concept includes three aspect in our baseline analysis (political competition, political participation, freedom of opinion), it consists of four aspects in this robustness check (political competition, political participation, freedom of opinion, judiciary independence).

Table C.40 Consequences of using different numerical forms — Threshold approach, alternative regime characteristics, OLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.026^{***}	0.007^{***}	0.008^{***}	0.009^{***}	0.012^{***}	0.011^{***}	0.011^{***}
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Equal. (p-val.)	–	0.000	0.001	0.001	0.007	0.003	0.005
	Panel B: Item-Response Approach						
Democracy	0.026^{***}	0.005^{***}	0.006^{***}	0.008^{***}	0.012^{***}	0.011^{***}	0.011^{***}
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Equal. (p-val.)	–	0.000	0.000	0.000	0.004	0.003	0.002
	Panel C: Multiplicative Approach						
Democracy	0.030^{***}	0.010^{***}	0.012^{***}	0.011^{***}	0.012^{***}	0.013^{***}	0.006^{***}
	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Equal. (p-val.)	–	0.000	0.001	0.001	0.001	0.002	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.031^{***}	0.010^{***}	0.012^{***}	0.009^{***}	0.012^{***}	0.011^{***}	0.013^{***}
	(0.006)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Equal. (p-val.)	–	0.000	0.001	0.000	0.001	0.001	0.002
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.026^{***}	0.010^{***}	0.012^{***}	0.009^{***}	0.012^{***}	0.011^{***}	0.013^{***}
	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Equal. (p-val.)	–	0.001	0.005	0.001	0.005	0.002	0.010

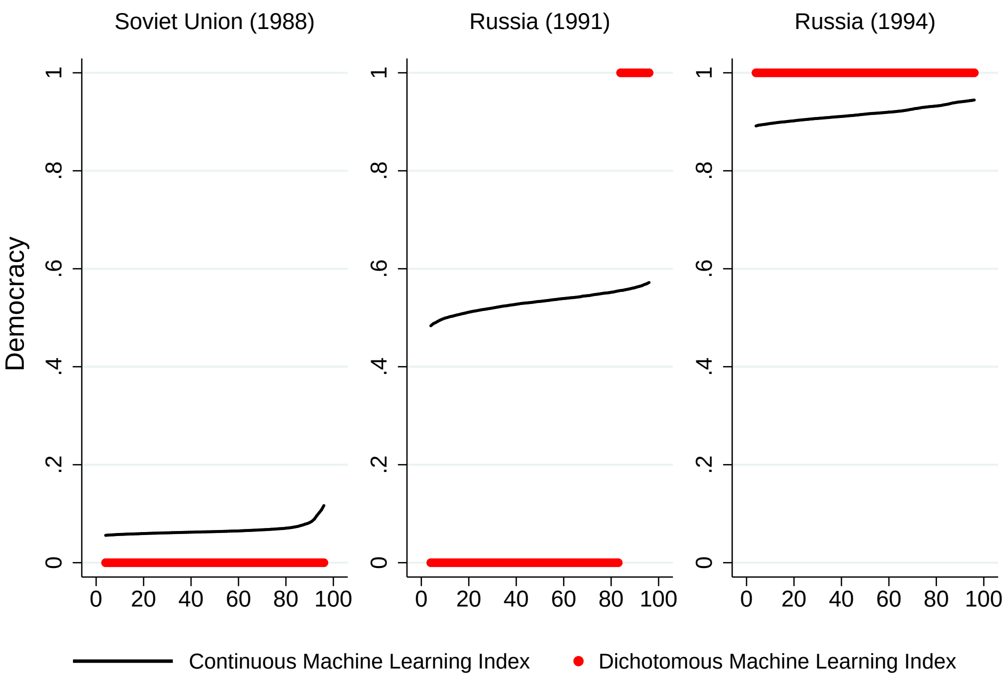
Notes: This table presents OLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The only difference compared to our baseline analysis (see Tables C.24 and C.26 is that we use the regime characteristics proposed by Teorell et al. (2019).

Table C.41 Consequences of using different numerical forms — Threshold approach, alternative regime characteristics, 2SLS estimates.

	Continuous	Threshold (0.3)	Threshold (0.4)	Threshold (0.5)	Threshold (0.6)	Threshold (0.7)	Threshold (0.8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Additive Approach						
Democracy	0.048***	0.055***	0.028***	0.025***	0.026***	0.028***	0.043***
	(0.009)	(0.013)	(0.006)	(0.005)	(0.005)	(0.005)	(0.009)
Equal. (p-val.)	–	0.462	0.022	0.008	0.014	0.022	0.533
	Panel B: Item-Response Approach						
Democracy	0.049***	0.056***	0.030***	0.025***	0.028***	0.037***	0.048**
	(0.010)	(0.015)	(0.007)	(0.005)	(0.006)	(0.007)	(0.010)
Equal. (p-val.)	–	0.480	0.051	0.013	0.029	0.191	0.888
	Panel C: Multiplicative Approach						
Democracy	0.057***	0.041***	0.045***	0.057***	0.067***	0.079***	0.118***
	(0.010)	(0.008)	(0.009)	(0.012)	(0.014)	(0.018)	(0.032)
Equal. (p-val.)	–	0.095	0.232	0.961	0.346	0.031	0.000
	Panel D: Additive/ Multiplicative Approach (Weigthed Average)						
Democracy	0.055***	0.028***	0.030***	0.035***	0.045***	0.063***	0.106***
	(0.009)	(0.005)	(0.005)	(0.006)	(0.008)	(0.013)	(0.029)
Equal. (p-val.)	–	0.004	0.008	0.034	0.281	0.384	0.000
	Panel E: Additive/ Multiplicative Approach (CD function)						
Democracy	0.046***	0.029***	0.030***	0.035***	0.044***	0.061***	0.098***
	(0.008)	(0.005)	(0.005)	(0.006)	(0.008)	(0.013)	(0.026)
Equal. (p-val.)	–	0.029	0.043	0.154	0.801	0.060	0.000

Notes: This table presents 2SLS estimates. The dependent variable is the log of GDP per capita. All regressions include four lags of the dependent variable, country fixed effects, and year fixed effects. The only difference between the seven columns is the democracy indices. Column 1 uses a continuous index, while Columns 2 – 7 use dichotomous indices. We construct these dichotomous indices by defining a threshold that must reach to be labeled as democratic. We indicate the threshold in the head of the table. The regional (jack-knifed) degree of democratization serves as the instrument for the domestic degree of democratization. We report the first-stage diagnostics proposed by Anderson and Rubin (1949) and Sanderson and Windmeijer (2016) to indicate the strength our instrumental variable. Standard errors clustered by country are reported in parentheses. We report results from a Wald test to show whether the estimates reported in Columns 2 – 7 are significantly different from the estimates reported in Column 1. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The regression results reported in Column 1 are not identical with the respective 2SLS estimates reported in Table C.16 because we only use the continuous index stated in the name of the panel to compute the regional degree of democratization. Our results do not change if we use the original instruments. The only difference compared to our baseline analysis (see Tables C.25 and C.27) is that we use the regime characteristics proposed by Teorell et al. (2019).

Figure C.1 Measurement uncertainty in the Machine Learning Indices



Notes: This figure presents confidence intervals to illustrate how measurement uncertainty differs between the continuous and dichotomous machine learning index. We observe that the measurement uncertainty in the dichotomous index is larger if a political transition takes place (see Russia, 1991). Prior and after a political transition (see Soviet Union, 1988, and Russia, 1994), the confidence intervals are greater for the continuous index.