Should we care (more) about data aggregation?

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

October 30, 2021

Abstract

Aggregation tools transform multidimensional data into indices. To investigate how the design of an aggregation process affects regression results, we build democracy indices that differ regarding their scale and aggregation function. Using the democracy-growth nexus as a testing ground, we illustrate that the choice of the aggregation procedure significantly affects OLS and 2SLS estimates since different methods produce systematically different index values for observations at the lower and upper end of the autocracy-democracy spectrum. We also illustrate that dichotomous measures produce significantly smaller OLS estimates than continuous measures due to lower discriminating power. Whether continuous and dichotomous indicators create different 2SLS estimates depends on their design. Because of the methodological similarities of democracy indicators and other social science indicators, we expect similar consequences for other empirical analyses.

Keywords: Aggregation, Data Transformation, Democracy, Economic Development, Indices, Machine Learning, Measurement of Democracy, Political Transitions, Scaling

JEL No.: C26, C43, O10, O43 P16, P48

Acknowledgements: We are grateful for the constructive comments from the editor (Florin O. Bilbiie), an associate editor, two anonymous referees, as well as 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]LMU Munich, ifo Center for Public Finance and Political Economy, Poschingerstrasse 5, 81679 Munich, Germany. E-Mail: Gruendler@ifo.de

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

1 Introduction

Democracy, globalization, gender inequality, and life satisfaction are just a few of the phenomena whose causes and consequences have been intensively studied in economics, sociology, and political science. A common feature of all these phenomena is that they are not directly observable and multidimensional. From a methodological perspective, this feature begs the question of how to measure complex social science phenomena. The common solution for this problem is to compile data for the different aspects of the phenomena and to choose an aggregation method that transforms the gathered data into an unidimensional index (Goertz, 2006, 2020).

Although indices exist for various social science phenomena[1] and have been frequently used in empirical studies, only very little is known about how the decisions that providers make during the creation process affect the results of regression analyses. The purpose of our paper is to fill this research gap. Our motivation is twofold. First, scholars have a lot of leeway when producing an index. We aim to support these researchers by illustrating the consequences of particular decisions. Unintended consequences are therefore less likely to occur. Second, related studies suggest that different indices produce different regression results (see e.g. Casper and Tufis, 2003, Cheibub et al., 2010, Gründler and Krieger, 2016). In this study, we highlight some reasons for these differences.

In social science, building an indicator is a three-step procedure. The initial step is to conceptualize the economic, political, or social phenomena of interest. The second step is to operationalize the components of this concept. The last step is to choose an aggregation rule that transforms the multidimensional raw data into an unidimensional index (Goertz, 2020, Hawken and Munck, 2013, Munck and Verkuilen, 2002). In each of these steps, several choices need to be made where each of these choices might have empirical consequences. The purpose of our project is to shed light on the consequences of choices made during the aggregation process.[2] We focus on these choices because the conceptual choices have often been discussed and are case-specific.[3] By contrast, the assumptions that are made when choosing an aggregation method are the same for all social science phenomena. The empirical consequences that result from replacing one

[1] Popular examples include the KOF Globalization Index, the indices of economic freedom by the Fraser Institute or the Heritage Foundation, the Worldwide Governance Indicators (WGI), the democracy indices by the Economist Intelligence Unit, Freedom House, or the Center of Systematic Peace, the Human Development Index, and the Gender Inequality Index of the United Nations Development Programme.

[2] Data aggregation requires two basic decisions (Hawken and Munck, 2013, Munck and Verkuilen, 2002): the first concerns the numerical form of the index, while the second decision is to specify the shape of the aggregation function. In our paper, we examine the consequences of both of these choices.

[3] For instance, Gutmann and Voigt (2018) and Voigt (2012) discuss how to conceptualize the rule of law, whereas De Haan and Sturm (2000) and Gwartney and Lawson (2003) describe the pros and cons of different concepts of economic freedom.

aggregation rule by another are thus likely to be similar for all phenomena in economics, political science, and sociology (Goertz, 2006, Wuttke et al., 2020).

We begin our analysis with a simulation study in which we use randomly generated data to provide a first impression of the empirical consequences that arise from using different aggregation techniques. Our simulation study suggests that the choice of the aggregation method notably affects the resulting index values, especially at the ends of the distribution. We also find that the results of OLS regressions change significantly if we replace one aggregation method by another.

To check whether the results of our simulation analysis reappear when using real data, we create measures of democracy that differ only in their numerical form (continuous vs. dichotomous) and their assumptions regarding the functional relationship between a set of observable regime characteristics and the level of democracy. We use the measurement of democracy as an expository example because there are various studies that controversially discuss the question of how to build measures of democracy (for surveys, see e.g. Boese, 2019, Gründler and Krieger, 2021, Munck and Verkuilen, 2002). With our choice, we therefore not only improve the general understanding of how aggregation procedures influence regression results, but also directly contribute to a longstanding debate in the political economy literature. Furthermore, we do not have to apply "artificial" aggregation tools when choosing democracy as a testing ground since the related literature includes a variety of different procedures (see e.g. Pemstein et al., 2010, Skaaning et al., 2015, Teorell et al., 2019). Finally, the causes and consequences of democratic transitions are at the very heart of the political economy literature. Since existing empirical studies often produce ambiguous results, we believe that it is worth examining whether this ambiguity can be partly explained by the fact that commonly used democracy indicators are designed with different data aggregation methods.

In the first part of our real data analysis, we compare six continuous indices. The only difference between these six indices is that they are computed with different data aggregation procedures. Our list includes: an additive approach as, for example, used by Marshall et al. (2019), the item-response approach proposed by Pemstein et al. (2010), a multiplicative approach (see e.g. Vanhanen, 2000), the Machine Learning approach developed by Gründler and Krieger (2016, 2021), and two approaches that combine an additive and a multiplicative measure (see e.g. Teorell et al., 2019). Consistent with the results of our initial simulation, we find in the real data analysis that the index values of regimes at the ends of the autocracy-democracy spectrum depend on the choice of the aggregation tool. A consequence of these differences is that our six measures differ notably in the extent to which they change after a regime transition. To examine the empirical consequences of using different aggregation methods, we run multiple regressions with each of the six indices. In most of the regressions, the level of economic development serves as our dependent variable. Alternative outcome variables are the average years of schooling and an expert-based measure of private property rights. Our findings suggest that the choice of the aggregation function has a statistically significant effect on the magnitude of OLS estimates. We also find that the differences in the estimates persist if we address endogeneity problems with a two-stage least squares approach. This finding is remarkable since many researchers believe that the choice of the composite measure is irrelevant if an appropriate instrumental variable exists. We argue that this presumption is unfounded and present a simple econometric model to substantiate our view. The basic argument of this model is that the aggregation functions produce non-classical measurement errors and their usage therefore leads to upward-biased estimates in both OLS and 2SLS regressions. Finally, we run a simulation to examine which of our six data aggregation methods creates the smallest bias in empirical analysis. This simulation suggests that the Machine Learning approach produce the smallest bias.

Another decision that leaves reseachers with great leeway is the scale of their index. For democracy and other social science indices, scholars usually either choose a continuous or a dichotomous scale. Since the related literature provides only very little information about how this decision influences regression results, the second part of our real data analysis studies whether decisions regarding the scale of an index also cause notable empirical consequences. To address this question, we need continuous and dichotomous democracy indices that are conceptually equivalent. We apply 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, 2021). The other method transforms a continuous indicator by defining a threshold value 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 argue that this difference arises since binary indices lack discriminating power. In instrumental variable regressions, the continuous and dichotomous Machine Learning index produce estimates that are not statistically significant from each other. When using the threshold approach, we also observe that the magnitude of the OLS estimate becomes smaller if 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 regimes to either the group of autocracies or democracies.

We are not aware of any study that examines how aggregation tools affect regression results. The studies that are closest to our analysis are Teorell et al. (2019) and Wuttke et al. (2020) who show how the distribution of index values depends on the choice of the aggregation function. We think that our work enhances these studies since indices are not only used for descriptive purposes but also often in regression analyses as dependent or explanatory variable. Our findings suggest that the choice of the data aggregation tool matters and thus urge researchers to be cautious when creating social science indices. Our paper also helps to anticipate the empirical consequences of the decisions that have to be made during the aggregation process and provides practical guidelines.

We also complement the literature on the measurement of democracy. Several studies suggest that the 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 explain where the differences in the regression results come from. In particular, Knutsen and Wig (2015) show that conceptual differences play a role. Elkins (2000) finds that the scale of the index has a significant effect on OLS estimates.[4]

Our paper also contributes to the literature that investigates how political transitions affect economic development. In line with Acemoglu et al. (2019) and Madsen et al. (2015), we find that transitions from autocracy to democracy fuel long-run growth. Our results also suggest that improved education and better economic institutions are two of the transmission channels. Finally, our paper implies that some of the variation in the estimated growth-promoting effects of democracy is caused by the use of different data aggregation methods.

We structure our paper as follows. Section 2 describes the framework that is commonly used to create indices in social science and presents our simulation study. Section 3 discusses different data aggregation methods and shows how changes in the aggregation function affects index values. Section 4 presents our empirical framework. Section 5 illustrates how the choice of the aggregation method shapes regression results. Section 6 compares the empirical performance of binary and continuous indicators. Section 7 concludes and provides practical guidelines.

2 Building indices in social science

2.1 General framework

The usual procedure for creating a social science index consists of three main steps (see e.g. Gwartney and Lawson, 2003, Hawken and Munck, 2013, Munck and Verkuilen, 2002). The first step conceptualizes the phenomena of interest. More specifically, in the first step of the creation process, scholars must address two

[4]Our analysis of the differences between continuous and dichotomous indices differs from Elkins' analysis in the following ways. First, Elkins (2000) does not use the Machine Learning tool. Second, Elkins (2000) only shows results from OLS regressions (which are consistent with our findings). Third, we present a greater number of robustness checks. Finally, we provide a detailed explanation for our results.

conceptual questions: (i) what are the *aspects* (*dimensions*) that are associated with the phenomena of interest, and (ii) how do they interact with each other. With regard to the first question, the literature distinguishes between narrow, realistic, and broad concepts (see e.g. O'Donnell, 2001, Voigt, 2012, Gutmann and Voigt, 2018). Theoretically, all types of concepts are equally valid because objective evaluation criteria do not exist (Guttman, 1994, Munck and Verkuilen, 2002). However, from an empirical point of view, narrow and broad concepts may create problems: while broad concepts are rather difficult to operationalize due to insufficient data availability and often cause conceptual overlaps, indices with narrow concepts typically lack sufficient discriminating power (Munck and Verkuilen, 2002, Voigt, 2012). Regarding the second key conceptual question, the literature presents two basic theories on how the single aspects of a particular phenomena interact with each other (Goertz, 2006, 2020, Teorell et al., 2019). The first theory assumes that each dimension constitutes a necessary condition. By contrast, the second theory suggests that all aspects are (partial) substitutes. From a theoretical perspective, neither of the two theories is superior because both of them have their strengths and weaknesses. Providers of indices should nonetheless pay attention to this matter because it affects the choice of the aggregation method.

The second step when building a social science index is to operationalize the aspects of the phenomena of interest. Put differently, for each of the aspects, a set of observable characteristics has to be found. The literature suggests some best practices for this task. For example, Munck and Verkuilen (2002) recommend using disaggregated raw data and compiling information from both objective and subjective data sources. In addition, each characteristic should be related to one dimension of the phenomena of interest.

The final step for each social scientist who aims to produce an index is to select an aggregation procedure that transforms the multidimensional raw data into an unidimensional measure. In formal terms, this means that we need to specify an aggregation function $\mathfrak{F}^c\colon \mathcal{X} \to \mathcal{D}$ that assigns an index value (Δ^c) to every concrete set of characteristics ($\mathbf{x}$):

$$\Delta^c := \mathfrak{F}^{\mathfrak{c}}(\mathbf{x}) = \mathfrak{F}^{\mathfrak{c}}(x_1, \ldots, x_m) \qquad \text{with} \qquad \mathbf{x} \in \mathcal{X}, \tag{1}$$

where c denotes the assumed concept of the phenomena of interest and m the number of characteristics compiled during the operationalization process. When computing (1), two crucial decisions must be made: (i) choosing the numerical form of the index and (ii) specifying the functional shape of the aggregation function. The latter decision is difficult since the true functional relationship of the characteristics and the phenomena of interest is unknown. In addition, the conceptual decision on how the different dimensions of the phenomena interact with each other only excludes a few of the potential aggregation functions. For instance, if all dimensions are assumed to be necessary conditions, only a few additive aggregation functions can immediately be considered as inappropriate. Consequently, providers of social science indices enjoy considerable leeway when designing their aggregation function. Munck and Verkuilen (2002) suggest that an unpleasant consequence of this leeway is that people often make simplistic and rather arbitrary functional assumptions. We share their views and thus believe that it is important to get a more profound understanding of the empirical consequences that result from these assumptions.

2.2 Do aggregation functions matter? A simulation example.

To give a first impression of the empirical consequences that might arise from applying different data aggregation procedures, we begin our analysis with a simulation that compares the performance of the two most common aggregation techniques for a set of randomly generated characteristics. More specifically, we first create four vectors (x_1, x_2, x_3, x_4). Each of these vectors consists of 10,000 entries ($x_{i,j}$) that are randomly chosen out of nine different realization ($x_{i,j} \in \{0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1\}$) with $Pr[x_{i,j} = 0] = Pr[x_{i,j} = 0.125] = \ldots = Pr[x_{ij} = 1]$.[5] Next, we use our four vectors to generate 10,000 pseudo observations. The four characteristics of pseudo observation n are ($x_{n,1}$, $x_{n,2}, x_{n,3}$, $x_{n,4}$). Finally, we use an additive and a multiplicative aggregation function to produce two different indices:

$$\Delta_i^{add} = \frac{1}{4} \cdot \sum_{j=1}^{4} x_{i,j} \qquad \text{and} \qquad \Delta_i^{multi} = \prod_{j=1}^{4} x_{i,j}^{0.25} \tag{2}$$

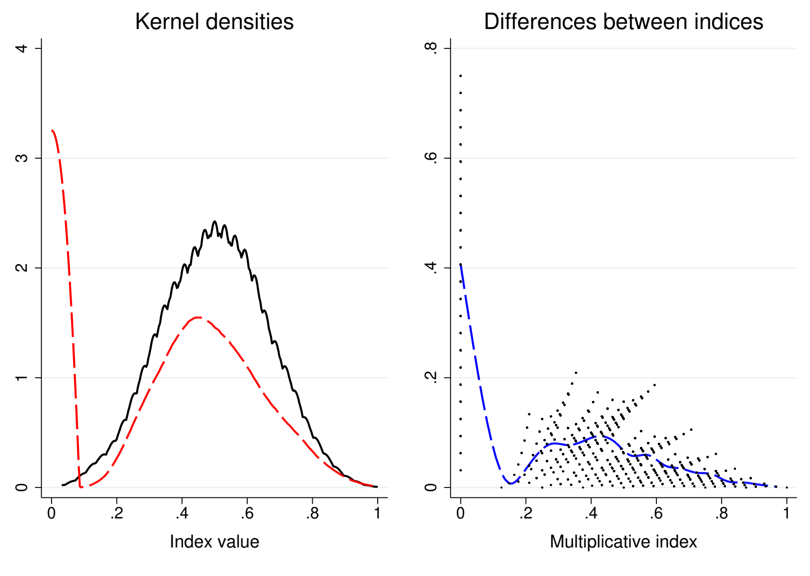
where i denotes the observation.

We proceed in three steps to compare the performance of the additive and multiplicative index. The first is to compute the correlation between the two indicators. We observe that that their pairwise correlation is positive (Pearson's correlation coefficient: 0.775; Spearman's rank correlation coefficient: 0.779). Next, we consider the results of kernel density estimations (see Figure 1, left graph). We find notable differences. In particular, we see that the density function is single-peaked for the additive index, whereas it is bimodal for the multiplicative index. The kernel functions also suggest that the differences between the two indices are most pronounced at the lower end of the distribution. To confirm this pattern, we finally compute the difference between the additive and the multiplicative indicator and illustrate how this difference depends on the index values of the multiplicative measure (see Figure 1, right panel).

From a practical perspective, an important question is whether the differences that we observe in the distributions have notable consequences when using the

[5] We choose a discrete rather than a continuous domain because this is much more common in practical applications.

Figure 1 Simulation example (distribution of index values)



Notes: The left figure shows the kernel density functions of the additive index (solid black line) and the multiplicative index (dashed red line). The right figure presents a scatter plot and an estimated non-linear function (dashed blue line) that shows how the difference between the additive and the multiplicative index varies within the spectrum.

indicators in an empirical analysis. To show that this might be the case, we assume that the true (in practice unobserved) index value is the mean of the additive and the multiplicative index:

$$\Delta_i^* = \frac{1}{2}\left(\Delta_i^{add} + \Delta_i^{multi}\right). \tag{3}$$

We also suppose that an economic outcome (Y) exists that is observable and determined by the following data-generation process:

$$Y = \beta \cdot \Delta_i^* + \varepsilon_i \quad \text{with} \quad \varepsilon \sim \mathcal{N}(0, 0.25) \quad \text{and} \quad \beta = 2. \tag{4}$$

Column 1 of Table 1 reports the results from a bivariate OLS regression in which we use the additive measure as the explanatory and Y as the dependent variable. We find that our point estimate is significantly larger than the true value of β. By contrast, if we apply the multiplicative index, we considerably underestimate the true parameter (see Column 2).

Taken together, our simple example suggests that index values, and especially those at the end of the distribution, depend on the choice of the aggregation function. We also observe that the results of a regression analysis can change considerably if we replace one aggregation method by another. Of course, at this stage of the analysis, our findings must be interpreted with caution because we only compare two aggregation procedures and use randomly generated data. In addition, we always assume that the regime characteristics are independent from

Table 1 Simulation example (regression results)

	(1)	(2)
Additive index	2.271***	
	(0.0178)	
Multiplicative index		1.446***
		(0.0103)
Regression technique	OLS	OLS
Observations	10,000	10,000
True parameter (β)	2.00	2.00

Notes: This table reports OLS estimates. The standard errors are presented in parentheses. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

each other. Put differently, a limitation of our simulation is that differences in the index values might only exist due to the particular data or be caused by the particular choice of aggregation methods. In addition, even if we find similar patterns in practical cases, it is unclear whether they cause the same empirical consequences as in our example. The remainder of this paper aims to address these concerns by considering a specific social science phenomena.

3 Measuring democracy

We illustrate the empirical consequences of using different aggregation methods based on the measurement of democracy. We focus on this phenomena because a large number of democracy indicators already exists. The providers of these indices use different aggregation methods and we thus do not have to consider "artificial" aggregation functions for our comparative analysis. Furthermore, the question of how democracy affects economic outcomes (such as long-run growth, inequality, and economic freedom) is at the heart of the comparative political economy literature. Existing empirical studies give rather differing answers to this question, which is why we believe that it is crucial to test whether divergent regression results can be explained with the use of different data aggregation methods.

3.1 Conceptualization

3.1.1 Dimensions of democracy

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

Munck and Verkuilen (2002) argue that theoretical discussions on the "correct" conceptualization of democracy are pointless since clear evaluation standards are lacking. We share this view. However, as explained in Section 2.1, narrow and broad concepts may create empirical problems (Bjørnskov and Rode, 2020, Munck and Verkuilen, 2002). Broad concepts are often difficult to operationalize because of insufficient data availability and are likely to overlap with other economic concepts, such as corruption, economic freedom, and social inequality. By contrast, narrow concepts usually have too little discriminating power. In our analysis, we therefore work with a realistic concept of democracy. This concept consists of three aspects: *political participation, political competition*, and *freedom of opinion.*

3.1.2 Interaction between the dimensions of democracy

The related literature presents two theories on how the individual dimensions of democracy interact with each other (Teorell et al., 2019). The first theory is that each aspect constitutes a necessary condition of democracy (Goertz, 2006, Boix et al., 2013). 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 aspects of democracy are (partial) substitutes (Treier and Jackman, 2008, Bollen, 1980, 1990). The justification for this approach is that all aspects correlate with each other and thus constitute a set of interchangeable "symptoms" of democracy (Teorell et al., 2019).

We believe that both theories have their merits and refrain from taking sides in this controversial debate. Below, we rather proceed with both of the theories. To study the consequences of using different aggregation methods, this neutral approach is of advantage since the decisions that we make at this stage of the creation process affect the choice of the aggregation method (Munck and Verkuilen, 2002). Put differently, if we opt for one of the two theories, we cannot compare the performance of all commonly used aggregated tools since most of them are not compatible with both theories.

3.2 Operationalization

We use ten regime characteristics that are available for a comprehensive sample of country-years to operationalize our three dimensions of democracy (see also Gründler and Krieger, 2021). To satisfy 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 peoples' right to elect their political rulers and representatives (Dahl, 1971). The suffrage might be limited, either through constitutional restrictions that exclude citizens because of their gender, race, or income, and by non-constitutional restrictions that result from material law, civil war, or repression. To capture the extent of constitutional restrictions, we use data from the Varieties of Democracy (V-Dem) database on the share of adult citizens with legally granted suffrage (Coppedge et al., 2019). Measuring non-constitutional disenfranchisement is more difficult because of insufficient data availability. We address this problem by collecting data on voter turnout and calculating the voter-population ratio (see also Vanhanen, 2000).[6] 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 disadvantage 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 information on only 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 is an expert-based index of party pluralism that discerns between five regime types.[7] The other four regime characteristics are based on objective data and reflect: (i) the share of votes not won by the strongest party/candidate,[8] (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.

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 discussion from the V-Dem database (Coppedge et al., 2019).

[6] We compile our data from a number of sources. A documentation of the collected data can be found here: https://www.ml-democracy-index.net/downloads/.

[7] 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. To create this measure, we use data from the V-Dem database, the database of the Inter-Parliamentary Union, and the election handbooks by Nohlen et al. (1999), Nohlen et al. (2001), Nohlen (2005), and Nohlen and Stöver (2010).

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

3.3 Aggregation

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

3.3.1 The additive approach

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

$$\Delta^{\text{add}} = \omega_1 \cdot x_1 + \ldots + \omega_m \cdot x_m \qquad \text{with} \qquad \sum_{j=1}^{m} \omega_j = 1, \tag{5}$$

where $\omega_j \in (0,1)$ is the weight assigned to regime characteristic x_j. The key reason for this choice is that additive aggregation procedures implicitly assume that there are no interactions between the regime characteristics. An additive aggregation rule therefore fits well together with the conceptual assumption of partial substitutability.

The main challenge when implementing (5) is to specify the weights (ω_j). A common approach is to assign the same weight to all regime characteristics (see e.g. the Polity index or the indices published by Freedom House). According to Munck and Verkuilen (2002), equal weighting is not appropriate for our data since the number of regime characteristics differs between our three aspects of democracy. Instead, we use a Principle Component Analysis (PCA) to determine the individual weights (see also Coppedge et al., 2008, Dreher, 2006, Gygli et al., 2019).[9] We report these weights in Table 2.

Treier and Jackman (2008) suggest that any index entails some degree of uncertainty and thus call for measures that reflect the extent to which an index suffers from measurement uncertainty. The additive approach fails to meet this requirement. Another major concern against additive aggregation procedures is that the decision on how a specific regime characteristic affects the degree of democratization cannot fully be grounded in theory and therefore might appear arbitrary. For example, when using (5), we assume 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 also achieve consistency

[9] We proceeded in two steps to create our weighting scheme. In the first step, we run a PCA for each of our three dimensions of democracy and computed additive sub-indices for all of them. In the second step, we applied the PCA to the three sub-indices and used its results to create our final index. In both steps, we only use the first component of the PCA to determine the weights.

Table 2 Weighting scheme of additive and multiplicative index.

Aspects of democracy	**Weight**
Political participation	
Suffrage	0.0515
Voter-Population ratio	0.0599
Voter turnout	0.0998
Political competition	
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 opinion	
Freedom of discussion (female)	0.1961
Freedom of discussion (male)	0.2009

Notes: This table presents 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) and Gygli et al. (2019).

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, (5) cannot be used to compute a dichotomous index. To address this issue, scholars who prefer a dichotomous over a continuous index often define a threshold value beyond which a regime can be considered as a democracy. Bogaards (2012) and Cheibub et al. (2010) criticize this procedure for two main reasons. First, this approach creates an inconsistency between theory and aggregation rule: while reaching a specific threshold constitutes a necessary condition, the conceptual assumptions associated with an additive index suggest that no necessary conditions exist. Second, any particular choice of a threshold value is arbitrary since it cannot be derived from theory. In Section 6, we show how regression results depend on this choice.

3.3.2 The item-response approach

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

$$x_{rj} = \Delta_r + \varepsilon_{rj} \quad \text{with} \quad \varepsilon_{rj} \sim \mathcal{N}(0, \sigma_j^2), \tag{6}$$

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.[10]

A practical challenge when using an item-response approach is that all regime characteristics need to have an ordinal scale with a finite number of categories (Treier and Jackman, 2008). Our regime characteristics do not meet this condition since 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).[11] Below, $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_m)$ will denote the ordinal version of our regime characteristics and $K = (K_1, \ldots, K_m)$ 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):

$$Pr\left(\hat{x}_{rj} = k \,|\, \Delta_r, \alpha_j, \sigma_j\right) = \mathcal{F}\left(\frac{\alpha_{j,k} - \Delta_r}{\sigma_j}\right) - \mathcal{F}\left(\frac{\alpha_{j,k-1} - \Delta_r}{\sigma_j}\right), \tag{7}$$

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:

$$L\left(\Delta, \alpha, \sigma\right) = \prod_{r=1}^{N} \prod_{j=1}^{m} \left[\mathcal{F}\left(\frac{\alpha_{j,\hat{x}_{rj}} - \Delta_r}{\sigma_j}\right) - \mathcal{F}\left(\frac{\alpha_{j,\hat{x}_{rj}-1} - \Delta_r}{\sigma_j}\right)\right], \tag{8}$$

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 estimates $\widehat{\Delta} = (\widehat{\Delta}_1, \ldots, \widehat{\Delta}_N)$ that we can use as democracy indices (Pemstein et al., 2010).

Compared to an additive aggregation method, 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,[12] 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)

[10]Three of our ten regime characteristics are discrete rather than continuous. A legitimate question is thus how the error terms can be independent from Δ_r as presumed by the item-response approach. Following Pemstein et al. (2010), we assume that our three discrete characteristics behave in a manner that is consistent with the idea that it represents a continuous underlying concept.

[11]The estimation results reported in Section 5 do not significantly change if we use alternative cutoffs.

[12]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) point out that this pattern is rather implausible because the measurement uncertainty should be largest for hybrid regimes and smallest at the extremes.

illustrate that one has to 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 create dichotomous indices (for a more detailed discussion of the threshold approach, see Section 3.3.1).

3.3.3 Multiplicative approach

Providers of democracy indices usually use a multiplicative aggregation method when assuming a concept in which a minimum of each aspect constitutes a necessary condition for democracy (Goertz, 2006):

$$\Delta^{\text{multi}} = x_1^{\omega_1} \cdot \ldots \cdot x_m^{\omega_m} \qquad \text{with} \qquad \omega_j \geq 0, \ \forall j \in \{1, \ldots, m\}. \tag{9}$$

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

Similar to an additive index, a main difficulty when creating a multiplicative indicator is to assign the weights to the regime characteristics. For the sake of consistency, we use the weighing scheme designed for the additive index in our baseline analysis (see Table 2). In several robustness checks, we show that our results hold if we use alternative weighting schemes (for further details, see Section 5.1). The multiplicative approach also shares a methodological problem with the additive approach because (9) does not create a measure that reflects the uncertainty of the indicator. A conceptual objection against (9) 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 approach that is consistent with such a concept of democracy (Goertz, 2006).

A key difference between additive and multiplicative aggregation tools is that multiplicative techniques can directly produce dichotomous indices. However, (9) creates a binary measure only if all regime characteristics are binary 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 3.3.1).

3.3.4 Combining additive and multiplicative indices

Teorell et al. (2019) illustrate that additive and multiplicative democracy indices have their greatest discriminatory power at opposite ends of the spectrum: while additive indices vary greatly for autocratic regimes and rather little for highly democratic regimes, multiplicative measures 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 the average of an additive and a multiplicative measure of democracy:

$$\Delta^{\text{av}} = \lambda \cdot \Delta^{\text{add}} + (1 - \lambda) \cdot \Delta^{\text{multi}} \tag{10}$$

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.[13]

Any concept assuming that specific aspects of democracy constitute necessary conditions is inconsistent with (10). The reason is that the level of democracy 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:

$$\Delta^{\text{cobb}} = \left(\Delta^{\text{add}}\right)^{\lambda} \cdot \left(\Delta^{\text{multi}}\right)^{1-\lambda} \quad \text{with} \quad \lambda \in (0, 1). \tag{11}$$

The aggregation procedures described by (10) and (11) share three weaknesses with the additive and multiplicative approach. First, none of them produces an indicator that reflects the degree of measurement uncertainty.[14] Second, the conceptual assumptions do not completely explain the shapes of the aggregation functions. Finally, the creation of a dichotomous democracy index requires the definition of an arbitrary threshold value (for a more detailed discussion of the threshold approach, see Section 3.3.1).

3.3.5 Machine Learning approach

In earlier studies, we proposed an aggregation procedure that is based on a Machine Learning technique for pattern recognition, known as Support Vector Machines (SVM) (see Gründler and Krieger, 2016, 2021). Our basic 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 Munck and Verkuilen, 2002, Cheibub et al., 2010). When using SVM, we can relax these assumptions and solve non-linear optimization problems to address the question of how to transform the regime characteristics into a measure of democracy. The downside of our method is that the shape of the aggregation function is not grounded in theory and lacks an explicit representation.

Since SVM is a supervised Machine Learning technique, its application requires a set of observations (henceforth: *priming data*) for which we observe both the

[13]The estimation results that we present in Section 5.1 do not significantly change if we assign the same weight to the additive and multiplicative index as proposed by Teorell et al. (2019).

[14]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.

input characteristics and the outcome variable (see e.g. Abe, 2005, Steinwart and Christmann, 2008). As outlined in great detail in Gründler and Krieger (2021), we proceed in two steps to meet 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 indicators by Pemstein et al. (2010) and Teorell et al. (2019) to identify regimes whose level of democracy is uncontroversial. In our basic version, we label an observation as a highly autocratic (democratic) regime if it belongs to the lower (upper) decile of either of the two indices.[15]

The data aggregation process of our Machine Learning approach includes four steps. In the first step, we randomly select country-year observations from the priming data to produce a training set $\mathcal{T}_\eta$. In the second step, we use a SVM tool and the training set $\mathcal{T}_\eta$ to estimate the aggregation function $\widehat{\mathfrak{F}}_\eta\colon [0,1]^m \to [0,1]$. The primary objective of this estimation is to find a function that aggregates the regime characteristics of the observations in the training set $\mathcal{T}_\eta$ such that the predicted outputs resemble the labels. To avoid overfitting, the objective function of the underlying optimization problem includes a penalization term that rewards smoothness (for more details, see Appendix A and Gründler and Krieger, 2016, 2021). 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:

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

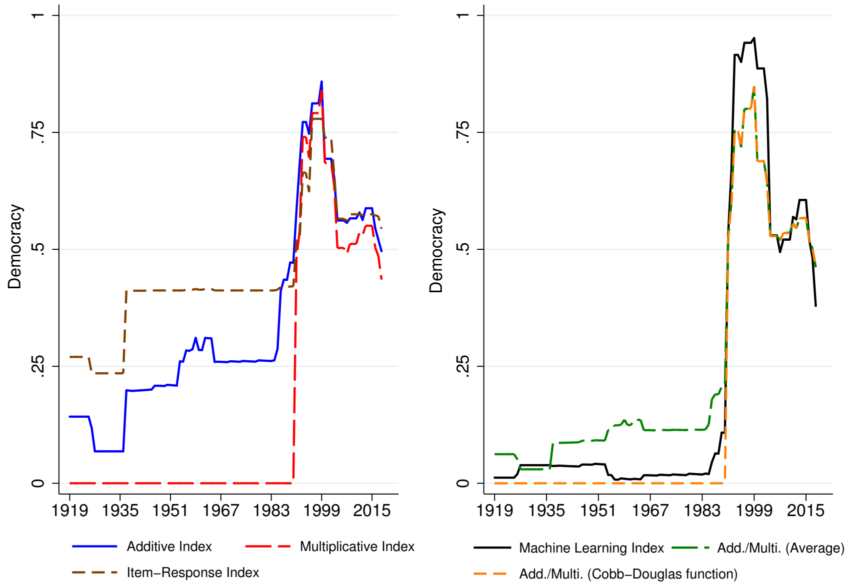
where i denotes the country and t the year. In the last step, we repeat steps $1 - 3$ for all iterations $\eta \in \{0, \ldots, \eta_{max}\}$. Our aggregation method thus creates a distribution of indices for each country-year observation. We use the median of each distribution as democracy indicator and other percentiles to reflect the degree of measurement uncertainty (see also Gründler and Krieger, 2021).[16]

The SVM toolbox includes classification and regression techniques (Abe, 2005, Steinwart and Christmann, 2008). We exploit these tools to produce (conceptually equivalent) continuous and dichotomous measures of democracy. Compared with the other aggregation procedures, our Machine Learning approach thus has the

[15] Gründler and Krieger (2021) show that the Machine Learning indices hardly change if we use alternative criteria. The regression results that we present in Sections 5 and 6 hold when applying other labeling procedures.

[16] 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. 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 must reflect the institutional heterogeneity among the autocratic and democratic regimes. Gründler and Krieger (2021) run various tests to illustrate that our priming data satisfies both conditions.

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.

advantage of producing binary indicators for all kinds of regime characteristics without requiring a manual definition of a threshold value. Another strength of our approach is that it creates measures of uncertainty for both the continuous and the dichotomous indices.

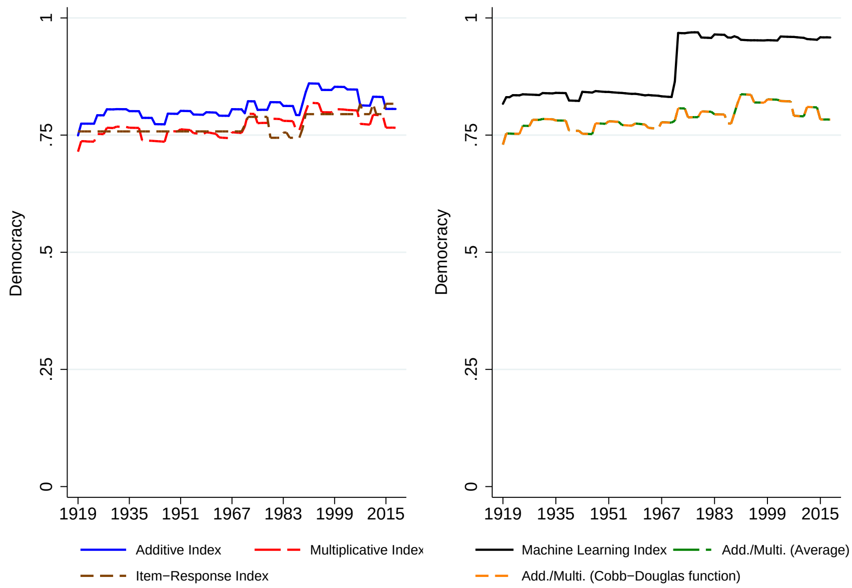
3.4 Comparing different aggregation methods

We now compare the performance of the six aggregation techniques that we described above. To this end, we proceed in two steps. In the first step, we present two example cases (Russia, Switzerland). These examples suggest that different aggregation methods produce different index values, especially for the regimes at ends of the autocracy-democracy spectrum. We also exemplify that indices therefore differ in the extent to which they change after a political transition. In the second step, we provide evidence, suggesting that our two examples illustrate a general rather than a case-specific pattern.

3.4.1 Example cases

In Figure 2, we present the level of democracy in the Soviet Union and the Russian Federation for different aggregation methods. We observe that all six measures indicate a distinct increase in the level of democracy 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,

Figure 3 Democracy in Switzerland (1919 – 2018).



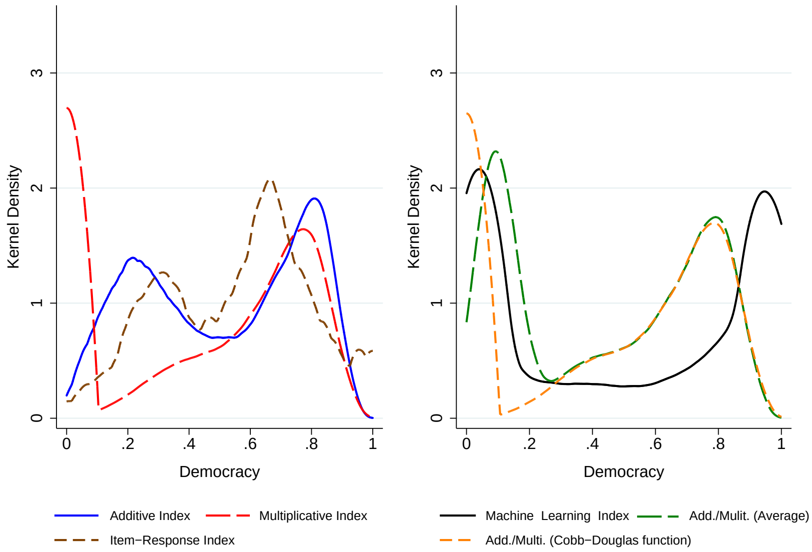
Notes: The figures show the level of democracy of Switzerland, depending on how we aggregate our ten regime characteristics.

whereas 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. Many expert reports confirm the plausibility of this decrease (see e.g. 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 function indicate the absence of democracy, the other three indicators suggest the existence of some democratic structures. These discrepancies exist because of the differences in the functional assumptions and the fact that non-competitive elections were held in the Soviet Union. For example, the additive measure is equal to 0 only if all the regime characteristics are equal to 0. By contrast, the multiplicative measure exceeds 0 only if all the 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 observe that the additive measure indicates a much higher level of democracy for the Soviet Union than the multiplicative measure. Consequently, the change of the additive indicator is less pronounced after the collapse of the Soviet Union.[17]

Our second example is Switzerland, a country that is widely acknowledged for its well-established democratic institutions (see Nohlen and Stöver, 2010). Our

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

Figure 4 Kernel density function



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

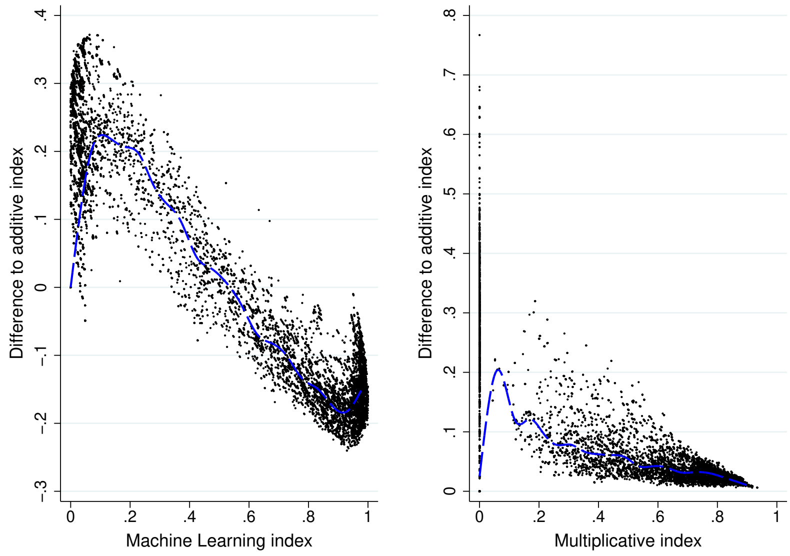
democracy indices reflect this institutional stability since we do not observe any significant decline in the degree of democratization (see Figure 3). Another fact for which Switzerland is well known is that it was the last European country that introduced female suffrage at the national level. Figure 3 shows that our indicators react differently to the enfranchisement of women in 1971: while the Machine Learning index indicates a notable increases in the level of democracy, the other five indicators change only marginally and still suggest a lack of democracy in Switzerland. The 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.

3.4.2 Generalization

The two examples presented above suggest that the choice of the aggregation function considerably affects the index values of highly democratic and highly autocratic regimes, and consequently show different reactions to an institutional transition. Of course, the patterns observed for Russia and Switzerland might suffer from low external validity. To ally this legitimate concern, we present a battery of additional tests in the remainder of this section.

In Figure 4, we present the results of kernel density estimations. We observe that all density functions are bimodal and have local maxima in the lower and upper part of the autocracy-democracy spectrum. However, the exact locations of the maxima differ notably 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

Figure 5 Differences between indices



Notes: The left (right) figure presents a scatter plot and an estimated non-linear function (dashed blue line) that shows how the differences between the additive and the Machine Learning (multiplicative) index varies within the spectrum.

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 highly autocratic and highly democratic regimes are, on average, less pronounced if we consider the additive rather than the Machine Learning index. Put differently, if a state experiences a transition from autocracy towards democracy, we can expect that the Machine Learning index indicates a larger change in the degree of democratization than the additive index.

The density functions shown in Figure 4 are consistent with our hypothesis that indices with different aggregation methods react differently to institutional transitions. However, when considering these functions, we cannot answer the question of where the different responses come from. We argue that the different reactions mainly exist because of differences for highly autocratic and highly democratic regimes. To substantiate our argument, Figure 5 shows two scatter plots. In the left panel, we illustrate how the difference between the Machine Learning and the additive index varies. In the right panel, we consider the multiplicative index rather than the Machine Leaning index. In line with the patterns that we observed in our example cases, we find that the additive and the Machine Learning indicator differ especially for regimes at the lower and upper end of the spectrum. The difference between the multiplicative index and the additive index is most pronounced for autocratic regimes. Interestingly, the pattern that we see in the right scatter plot is not only consistent with the

Table 3 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 (12). The figures show how much a democracy index change, on average, when Bjørnskov and Rode (2020), Boix et al. (2013), or Papaioannou and Siourounis (2008) indicate a political transition from autocracy to democracy (or vice versa).

results of our two example cases but also with the results of our simulation example (see Section 2.2).

Another way to illustrate that our six indices react differently to political transitions is to exploit the dichotomous indicators of Bjørnskov and Rode (2020), Boix et al. (2013), and Papaioannou and Siourounis (2008) and to create measures that reflect the average change in the degree of democratization during a transition:

$$\Theta_k^j = \frac{1}{|\mathcal{S}_k|} \sum_{(i,t) \in \mathcal{S}_k} |\, \Delta_{i,t}^j - \Delta_{i,t-1}^j \,| \tag{12}$$

where $\mathcal{S}_k$ is the set of regime changes either indicated by Bjørnskov and Rode (2020), 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 3. In line with the patterns observed in Figure 4, we find notable differences between our continuous indicators. In particular, we observe that the Machine Learning indicator shows the greatest response to a political transition, while the additive and the item-response indicator change relatively little.

4 Empirical framework

Having established that the choice of the aggregation tool influences the index values of highly autocratic/democratic regimes and thus the extent to which a democracy index reacts to a political transition, we now examine whether these different behaviors have empirical consequences. To address this question, we present 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 (see 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 this ambiguity are still unclear, we believe that improving the understanding of the role of the democracy index is a useful contribution to this literature. Second, the literature on the effect of democracy on economic development proposes various identification strategies. We exploit this variety to show how the consequences of using different aggregation tools depend on the applied regression technique.

According to Acemoglu et al. (2019), there exist three main endogeneity issues that complicate an analysis of the effect of democracy on growth. First of all, autocratic regimes differ from democratic regimes in non-observable factors that also affect growth. Second, causality might run from economic development to democracy.[18] Finally, democratization is often preceded by a temporal decline in GDP per capita. To address these endogeneity problems, most scholars estimate the following dynamic fixed effect model (see e.g. Acemoglu et al., 2019):

$$Y_{i,t} = \sum_{l=1}^{L} \beta_l \cdot Y_{i,t-l} + \gamma \cdot D_{i,t} + \xi_i + \eta_t + \varepsilon_{i,t} \tag{13}$$

where D denotes the level of democracy 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.[19]

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 conditional on our two sets of fixed effects and other covariates. Since this condition is unlikely to be satisfied because of omitted time-varying factors, various recent studies use a two-stage least squares (2SLS) approach in which the average level of democracy in the neighboring countries serves as the instrument for the domestic level of democracy (see e.g. Acemoglu et al., 2019, Dorsch and Maarek, 2019, Persson and Tabellini, 2009):

$$D_{i,t} = \sum_{l=1}^{L} \delta_l \cdot Y_{i,t-l} + \alpha \cdot Z_{i,t} + \zeta_i + \tau_t + \iota_{i,t} \tag{14}$$

with

$$Z_{i,t} = \frac{1}{|\mathcal{R}|} \sum_{j \in \mathcal{R}} D_{j,t} \quad \text{and} \quad \mathcal{R} = \{j : j \neq i, r_j = r_i\} \tag{15}$$

where r_i denotes the region in which country i is located.[20] The motivation for this identification strategy is that transitions from autocracy to democracy (and vice versa) often occur in regional waves (Huntington, 1993, Teorell, 2010).

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

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

[20] In our baseline analysis, we use the classification of the United Nations to divide the world into 19 regions. Our regression results do hardly change if we use other classification schemes.

The 2SLS approach creates valid estimates for the effect of democracy on economic growth if two assumptions hold (Angrist and Pischke, 2009): first, the regional and the domestic degree of democratization 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 second assumption might be violated. For example, changes in the regional degree of democratization might have an effect on the regional level of political stability, which in turn affects domestic prices, investments, and trade flows. Acemoglu et al. (2019) and Dorsch and Maarek (2019) present a battery of tests that allay concerns regarding the validity of the exclusion restriction.

5 Differences in the shape of the aggregation function

5.1 Estimation results

5.1.1 OLS estimates

In Table 4, we present OLS results from estimating Equation (13) with an unbalanced panel that covers 163 countries over the period from 1919 to 2016. The only difference between the six columns is that we applied different data aggregation tools to create the democracy index. In all six regressions, we add four lags of the dependent variable to our model. As common in the related literature, we cluster the standard errors at the country level.

Column 1 uses the Machine Learning index. In line with some 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 the per-capita GDP by 1.7 percent per year.[21] 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).[22]

In Column 2, we use the additive index. We observe that the change in the

[21]Compared to the OLS results reported by Papaioannou and Siourounis (2008) and Acemoglu et al. (2019), we find a larger estimate. In Section 6, 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 4.

[22]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 4 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
Wald test (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: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

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 instead of the additive approach (see Column 3).

Column 4 presents the result for the multiplicative indicator. Compared to the Machine Learning indicator, the OLS estimate of the effect of democracy on economic development increases significantly, but to a lesser extent than the additive measure. The gap between the regression coefficients produced by the additive index and the multiplicative index is statistically significant at the 5 percent level (p-value: 0.012) and is consistent with the pattern that we observe in our simulation example (see Section 2.2).

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 indicator 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 if 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 4 shows that the choice of the aggregation tool considerably affects the results of OLS regressions in a significant manner. We observe that the size of the estimated effect of democracy on economic growth almost doubles if we replace the aggregation technique that produces the smallest regression coefficient (see Column 1) with the aggregation method that produce the largest regression coefficient (see Column 3). In the next section, we examine whether the differences in the estimates disappear when using a 2SLS approach.

5.1.2 2SLS estimates

To investigate the consequences of using different aggregation methods for the results of 2SLS regressions, we would like to use an instrumental variable that does not depend on the design of the aggregation procedure. Unfortunately, the literature does not provide such an instrument. We thus need to 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:[23]

$$Z_{i,t} = \frac{1}{6} \cdot \left(Z_{i,t}^{ML} + Z_{i,t}^{Add} + Z_{i,t}^{IR} + Z_{i,t}^{Multi} + Z_{i,t}^{AM_{Av}} + Z_{i,t}^{AM_{CD}} \right). \tag{16}$$

Table 5 shows 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 4, 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 e.g. 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 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 if we use the item-response approach to create a continuous measure of democracy.

Column 4 presents the regression results for the multiplicative measure. We observe that this indicator produces first- and second-stage estimates that lie between the estimates of the Machine Learning index and the additive index. A notable difference compared to the OLS estimates reported in Table 4 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

[23]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 the second-stage estimates hardly change compared to the estimates reported in Table 5. The differences in the first-stage estimates are smaller than in the baseline analysis. In our main analysis, we focus on the case in which the instrument is the same in all specifications. We think this is the more appropriate procedure because we aim to illustrate the consequence of using different aggregation methods in a general manner and not only for jack-knifed instruments.

Table 5 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)
Wald test (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)
Wald test (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: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

the multiplicative index. The main 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 5 illustrates that a change in the aggregation method can create significant changes in the results of 2SLS regressions. This finding is notable because many scholars who examine the consequences of political transitions believe that the choice of the democracy index is irrelevant if one applies an instrumental variable approach.[24] In Section 5.2, we explain why this

[24]For instance, Acemoglu et al. (2019) whose IV approach is very similar to our IV approach claim that: "Our IV strategy also alleviates concerns related to measurement error in our

logic does not necessarily apply. 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 little external validity.

5.1.3 Additional results

Many economists argue that annual data is not appropriate for examining 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 illustrate that the choice of the aggregation method also affects the results of OLS and 2SLS regressions if we use five-year averaged data rather than annual data. However, the differences in the regression coefficients are slightly less pronounced than in our baseline analysis.

In Appendix Table C.5 and C.6, we extend our regression models by control variables that are available for the entire sample period (population growth, civil conflict, rule of law).[25] The results show that the inclusion of these covariates has only a small impact on the consequences of using different data aggregation methods.

Appendix Tables C.7 – C.10 presents results from regressions in which the average years of schooling and a subjective measure of private property rights serve as the outcome variables.[26] In line with related studies, we observe that both the education level and the quality of the economic institutions increases in the degree of democratization (see e.g. De Haan and Sturm, 2003, Harding and Stasavage, 2013). We also find that the differences between our six measures of democracy persist.

Another potential concern is that we strategically selected our ten regime characteristics and that the differences between the data aggregation methods disappear if we use alternative regime characteristics. To allay this legitimate concern, we repeat our baseline analysis with the regime characteristics used by Teorell et al. (2019). In Appendix Tables C.15 and C.16, we illustrate that our results hold if we use an alternative set of regime characteristics.

Aidt and Eterovic (2011) provide evidence suggesting that different aspects of democracy have different economic effects. A reason for why the choice of the aggregation method affects regression results might thus be that different data aggregation methods put different weights on different aspects of democracy. To

measure of democracy, [...] ."

[25] 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.

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

check whether this reasoning applies, we create indices that only consider the aspect of political competition. Appendix Tables C.11 and C.12 show that the choice of the aggregation tool continues to matter for the regression results. In Appendix Tables C.13 and C.14, we illustrate that the differences between the indicators also do not disappear if we use *judiciary independence* as a fourth aspect of democracy. We thus doubt that conceptual issues can serve as an explanation for our main findings.

Finally, we examine whether our results are driven by the weighting schemes that we use to create our additive and multiplicative index. In our baseline analysis, we use a PCA. A potential objection against this approach might be that the PCA weights are estimated, which in turn might create biases in our regression results. We do not think that this is the case because our findings remain virtually unchanged if we weight all regime characteristics equally (see Appendix Tables C.17 and C.18).

In sum, the results of our robustness checks confirm that the choice of the aggregation method significantly affects the size of OLS and 2SLS estimates. Our results also show that the rank order of the estimates is fairly robust. In all regressions, we observe 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 combining the additive and the multiplicative index, we find that using a Cobb-Douglas function leads to smaller estimates than taking a weighted average.

5.2 Explanation

In Section 3.4, we illustrate that the choice of the aggregation rule significantly influences the index values of the regimes at the upper and the lower end of the autocracy-democracy spectrum. As a consequence, we see that indices differ systematically in their reactions to political transitions. In this section, we will present a stylized econometric model to show that these systematic differences cause differences in OLS and 2SLS estimates that are in line with the patterns observed in the previous section.

OLS estimates

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

$$Y_i = \alpha + \beta \cdot D_i + \varepsilon_i \tag{17}$$

where $\alpha, \beta > 0$ denote 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 indices and suppose that the first index (Δ^1) indicates a lower (higher) degree of democratization than the second (Δ^2) for regimes with a low (high) level of democracy:

$$\Delta_j^2 \;=\; \Delta_j^1 \;+\; \mathcal{E}\big(D_j\big) \quad \text{for} \quad j \in \{low,\, high\} \tag{18}$$

with

$$\mathcal{E}\big(D_j\big) \;=\; \begin{cases} -\,\eta & \text{for} \quad D_j \;=\; D_{high} \\ \;\gamma & \text{for} \quad D_j \;=\; D_{low} \end{cases} \quad \text{with} \quad \eta > 0 \quad \text{and} \quad \gamma > 0.^{27} \tag{19}$$

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):

$$\widehat{\beta}_{ols}^{k} \;=\; \frac{\text{cov}\left(Y,\, \Delta^k\right)}{\text{var}\left(\Delta^k\right)} \;=\; \frac{m \cdot \sum_{i=m+1}^{n}(Y_i - \bar{Y}) - (n-m) \cdot \sum_{i=1}^{m}(Y_i - \bar{Y})}{(\Delta_{high}^k - \Delta_{low}^k) \cdot m \cdot (n-m)} \tag{20}$$

where $k \in \{1,\, 2\}$ indicates whether we apply the first or the second democracy indicator. Equation (20) shows that the magnitude of the OLS estimate increases when the difference between Δ_{high}^k and Δ_{low}^k decreases. The second indicator thus produces larger OLS estimates than the first index:

$$\Delta_{high}^2 \,-\, \Delta_{low}^2 \;<\; \Delta_{high}^1 \,-\, \Delta_{low}^1 \quad \Rightarrow \quad \widehat{\beta}_{ols}^{2} \;>\; \widehat{\beta}_{ols}^{1}. \tag{21}$$

Consistent with this theoretical prediction, we find that the Machine Learning index creates the smallest OLS estimates (see Table 4) and changes most if a political transition takes place (see Table 3). By contrast, the additive and the item-response index indicate small changes and produce relatively large OLS estimates. Our model also fits well together with the patterns that we observe for two indices that we create by combining the additive and the multiplicative indicator.

2SLS estimates

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

[27] Below, we only consider cases in which $d_{2,high} \;=\; \mathrm{d}_{high} \,-\, \eta \;>\; \mathrm{d}_{low} \,+\, \gamma \;=\; d_{2,low}$ because we believe that this is the most relastic one.

of democracy. If we use Z as an instrumental variable in a 2SLS regression, we obtain the second-stage estimates:

$$\widehat{\beta}_{iv}^{k} = \frac{\text{cov}\left(Y, Z\right)}{\text{cov}\left(\Delta^{k}, Z\right)} = \frac{\widehat{\delta}_{ols}}{\widehat{\rho}_{ols}^{k}}, \tag{22}$$

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

$$\Delta^{k} = \pi + \rho \cdot Z_i + \xi_i \quad \text{with} \quad \rho > 0, \tag{23}$$

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

$$\text{Y}_i = \zeta + \delta \cdot \text{Z}_i + \iota_i \quad \text{with} \quad \delta > 0. \tag{24}$$

From

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

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:

$$\Delta_{high}^{2} - \Delta_{low}^{2} < \Delta_{high}^{1} - \Delta_{low}^{1} \quad \Rightarrow \quad \widehat{\rho}_{ols}^{2} < \widehat{\rho}_{ols}^{1} \quad \Rightarrow \quad \widehat{\beta}_{iv}^{2} > \widehat{\beta}_{iv}^{2}. \tag{25}$$

When considering the results in Tables 3 and 5, we observe that the predictions of our model match the patterns that we find in the data. In particular, the indices that indicate relatively small changes after a political transition create relatively small first- and relatively large second-stage estimates. The opposite is the case for the indices that change more extensively if a political transition occurs.[28]

5.3 Evaluation

The results reported in Section 5.1 illustrate that the choice of the aggregation function influences the estimates produced in OLS and 2SLS regressions. A key

[28] A concern might be that our theoretical model is much simpler than our actual regression model since it neither includes fixed effects nor lagged dependent variables. To address this concern, we check whether the differences that we observe between our six indicators persist if we estimate bivariate regression models. Appendix Tables C.19 and C.20 show that this is indeed the case.

question that leaves open from this real data analysis is which of the methods produces the regression coefficients that are closest to the true effects. Below, we address this question with a simulation analysis.

The starting point of our analysis is the observation that the choice of the aggregation method mainly affects the index values of regimes at the upper and lower end of the autocracy-democracy spectrum (for details, see Section 3.4). The differences in the regression results that we find when replacing one aggregation tool by another are therefore likely to be caused by the differential ratings of highly autocratic and highly democratic countries. Consequently, a natural first step when evaluating the performance of different aggregation tools is to study which method produces the best indices for these regimes. The main challenge when addressing this question is that the true levels of democracy cannot be observed. Our solution for this problem is to make a few assumptions about the relationship between the regime characteristics and the level of democracy. Based on these assumptions, we then derive the true levels of some regimes. In a final step, we compare these levels with the index values that we obtain when using our six aggregation methods.

Among the very few aspects on which almost all social scientists agree when discussing the question of how to define democracy is that a country where no political competition exists is non-democratic (see e.g. O'Donnell, 2001, Przeworski, 1991, Schumpeter, 1942). We conclude from this consensus that the true level of democracy is observable for regimes without political competition ($\Delta = 0$). From our perspective, this particular type of autocracies thus provides the opportunity to evaluate the performance of different aggregation procedures. To facilitate this analysis, we create 1,000 pseudo regimes whose characteristics imply the absences of political competition (for details on the data generation process, see Appendix B.1). We then apply our aggregation tools to the characteristics of our pseudo regimes and thereby calculate six democracy measures for each regime. For each aggregation method, we finally report the average level of democracy and report this mean in the first row of Table 6. We find that the multiplicative approach and the approach that combines an additive and a multiplicative index with a Cobb-Douglas function produce indices that correctly reflect the true value. When using the Machine Learning approach, we obtain measures that are, on average, slightly too high for regimes in which political competition does not exist. For the other aggregation methods, we observe that the index values are considerably larger than 0.

In a similar way as for the highly autocratic regimes, we create 1,000 pseudo regimes to study which aggregation tool performs best at the upper end of the spectrum (for details on the data generation process, see Appendix B.2). The results of this analysis are shown in the second row of Table 6. We find that the indices produced by the Machine Learning approach exceed, on average, the

Table 6 Comparing the performance of different aggregation methods (simulation, part I).

True value (pseudo regimes)	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta = 0$	0.023	0.121	0.290	0.000	0.066	0.000
$\Delta = 1$	0.955	0.903	0.923	0.888	0.896	0.896

Notes: This table shows the average levels of democracy that we obtain when applying our aggregation methods to our two sets of pseudo regimes. The true values are reported in the first column.

indices produced by all other data aggregation methods. The Machine Learning indices are thus closest to the true values. Among the five other methods, the only notable difference is that the item-responds approach creates slightly large index values.

The findings reported in Table 6 suggest that the Machine Learning indicator indicates, on average, the greatest change in the degree of democratization if a country moves from autocracy to democracy. The additive and the item-response measure show, by contrast, the least pronounced reaction to political transitions. From our perspective, this pattern is remarkable for two main reasons. First, it is consistent with the results that we find in the real data analysis (see Section 3.4). This consistency is crucial since we could otherwise hardly use the results of our simulation analysis to give a sound answer to the question of which data aggregation method performs best in real data analyses. Second, the theoretical framework presented in Section 5.2 suggests that the extent to which an index reacts to political transitions determines the size of the regression coefficient. In particular, our stylized model predicts that the bias in the regression coefficient increases in the extent to which an index underestimates the true change in the degree of democratization. Using the Machine Learning (item-response) index should thus produce the smallest (largest) bias in empirical analyses. To verify this prediction, we first generate outcome variables for our 2,000 pseudo regimes. The data generation process is:

$$Y \;=\; \beta \cdot \Delta_i^* \;+\; \varepsilon_i \qquad \text{with} \qquad \varepsilon \sim \mathcal{N}(0,\, 0.1) \qquad \text{and} \qquad \beta \;=\; 2. \tag{26}$$

Afterwards, we run a OLS regression with each of our six indicators. The results of these regressions are shown in Table 7 and are in line with the predictions of our stylized model. Reassuringly, the order of the six estimates is very similar to the order that we observe in our real data analysis (see Section 5.1).

Taken together, the results of our simulation analysis imply that the Machine Learning approach outperforms the other aggregation tools. The reason for this result is that the Machine Learning index performs relatively well at both ends of the autocracy-democracy spectrum. The second best aggregation methods are the multiplicative approach and the approach that combines an additive and a multiplicative index with a Cobb-Douglas function. A potential concern regarding

Table 7 Comparing the performance of different aggregation methods (simulation, part II).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	2.137***	2.522***	2.935***	2.245***	2.395***	2.226***
	(0.0056)	(0.0088)	(0.0194)	(0.0056)	(0.0067)	(0.0055)
Observations	2,000	2,000	2,000	2,000	2,000	2,000
True parameter	$\beta = 2$	$\beta = 2$	$\beta = 2$	$\beta = 2$	$\beta = 2$	$\beta = 2$

Notes: This table reports results from OLS regressions. The standard errors are presented in parentheses. The following notation is used to highlight coefficients that are significantly different from zero: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

our simulation analysis is that we focus on very specific regimes when comparing the performance of different aggregation tools. We are well aware of this problem and therefore recommend to interpret the results of our simulation with caution. However, we also believe that focusing on specific regimes is necessary since we would otherwise need very strong assumptions about the relationship between the regime characteristics and the level of democracy to derive the true values. From our perspective, making such assumptions is a bad idea because they have been heavily criticized in the related literature (see e.g. Cheibub et al., 2010, Munck and Verkuilen, 2002).

6 Differences in the numerical form

To investigate 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 measures that are conceptually equivalent. We will 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 measure 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.

6.1 Machine Learning Approach

A feature of the Machine Learning approach is that it does not need manual interventions to create (conceptually equivalent) binary and continuous measures. This feature exists since the SVM toolbox includes classification and regression methods that operate in a similar manner (for details, see Appendix A). We exploit this methodological flexibility to provide novel evidence on the question of how a change from a continuous to a dichotomous measure affects regressions results.

Table 8 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
Wald test (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: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

6.1.1 Estimation results

Columns 1 and 3 of Table 8 show that the OLS estimate of the effect of democracy on economic growth decreases from 1.7 percent to 1.0 percent if we apply the dichotomous rather than the continuous Machine Learning index. This decline is statistically significant at conventional levels and confirms Elkins (2000) who finds that continuous measures produce larger estimates than dichotomous measures. However, 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 suggests a smaller effect of transitions towards democracy than the continuous index. However, the point estimates do not differ in a statistically significant manner (p-value: 0.373). Table 8 thus suggests that the continuous and the dichotomous index behave differently in OLS regressions and similarly in 2SLS regressions.

To illustrate the external validity of the findings reported in Table 8, we run the same robustness checks as in Section 5.1.3. Appendix Tables C.21 – C.27 show the results. Three findings are especially notable. First, the estimate produced by the continuous indicator always exceeds the estimate produced by the dichotomous 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.

6.1.2 Explanation

For two reasons, we argue that differences in the discriminating power explain why the dichotomous Machine Learning index produces smaller OLS estimates than the continuous index. First, due to their low distinctiveness, dichotomous indices cannot reflect that political transitions often take place gradually.[29] Put differently, the year in which a binary measures switches from 0 to 1 (or vice versa) is typically not the same as the year in which a transition starts. The immediate economic effects of the events that initiate a transition process are therefore attributed to the old regime. As a consequence, we underestimate the growth effects of the transition when using a dichotomous index and the fixed effect approach because this approach only compares the economic performance before and after the year in which the binary index changes.[30] By contrast, a continuous index already reflects the first steps towards democracy. The fixed effect approach can therefore relate the economic consequences of the initiating reforms to a change in the level of democracy.

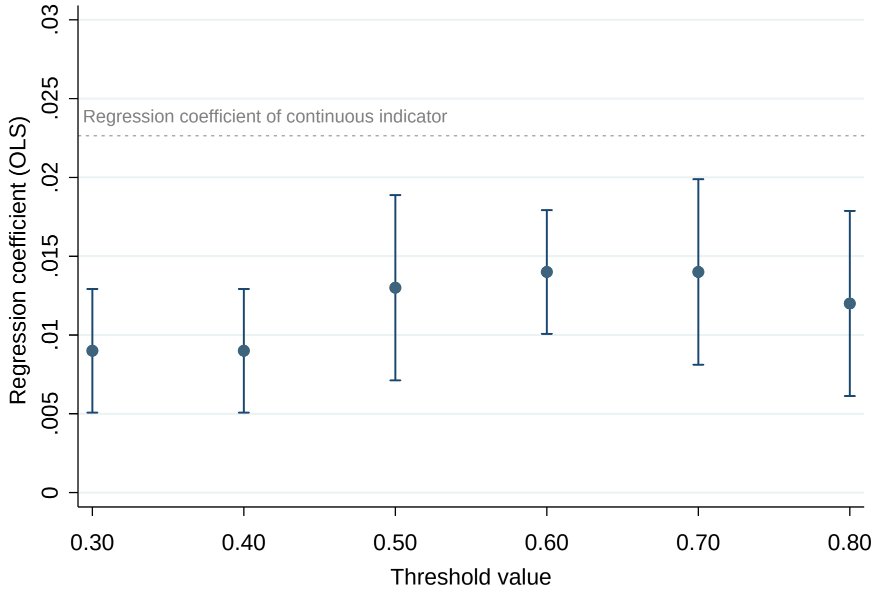
The second reason for why using the binary measure leads to smaller OLS estimates is that this index cannot differentiate between completely and partly democratic regimes. A partly democratic regime is thus either classified as an autocratic or democratic regime. Put differently, if a partly democratic regime replaces an autocratic regime, the dichotomous index either does not react or changes in the same manner as in the case when a regime switches from autocracy to full democracy. In a fixed effect regression, both issues cause a decrease in the estimates since partial transitions towards democracy enhance economic performance, but to a lesser extent than full transitions.

An important final question is why the differences in the regression results produced by the continuous and dichotomous Machine Learning index do not persist if we use the instrumental variable approach. The reason is that the measures of democracy predicted by the first-stage regressions and used in the second-stage regressions are continuous indices regardless of whether we exploit the continuous or the dichotomous Machine Learning indicator as the dependent variable in the first-stage regression. In our case, this has the effect that the mean absolute deviation between the predicted levels of democracy (0.112) is smaller than the actual difference between the continuous and the dichotomous

[29] Figure C.1 illustrates this issue based on the Russian indices. While the continuous index increases stepwise between the late-1980s and early-1990s, the dichotomous Machine Learning index jumps suddenly from 0 to 1 in 1992.

[30] The fixed effect approach under- rather than overestimates the growth effects of a political transition when using a binary measure because the first reforms that an autocratic regime undertakes when moving towards democracy already fuel growth (e.g. by attracting foreign investors), whereas growth already decreases if an initially democratic government becomes increasingly authoritarian.

Figure 6 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.28). 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.

index (0.033). As a consequence, we observe that the differences in the point estimates disappear when using our instrumental variable approach.

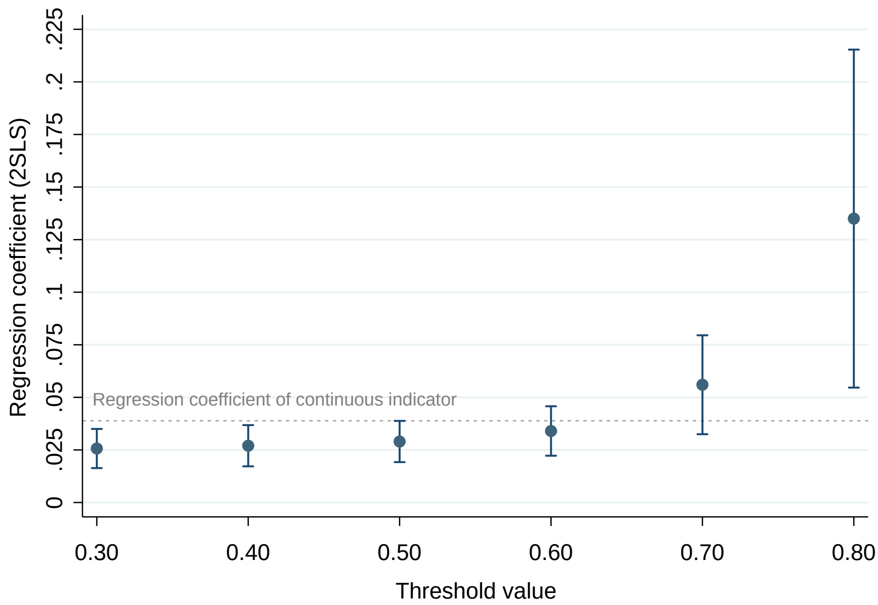
6.2 Defining threshold values

A second way to create a dichotomous measure of democracy is to select a continuous indicator and to choose a threshold up to which a regime can be classified as democratic. This method has been frequently criticized especially because the level of the threshold value is arbitrary (see e.g. Bogaards, 2012).[31] From an empirical point of view, this arbitrariness is worrisome if regression results significantly react to changes in the threshold. In this section, we test whether this problem is of practical relevance.

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 6 and 7 present the results of these regressions when using the multiplicative measure (for the tabulated results, see Appendix Tables C.28 and C.29). In both figures, the dashed line reflects the regression coefficient produced by the multiplicative indicator, while the dots show the

[31]Despite this harsh critique, this approach is still frequently applied (see e.g. Acemoglu et al., 2019).

Figure 7 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.29). 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.

estimates of the dichotomous indices. At the horizontal axes, we present the threshold that a regime must reach to be labeled as democratic ($\Delta = 1$). The vertical lines show the 95 percent confidence intervals associated with the point estimates.

We find that the choice of the threshold value has only a small effect on the magnitude of the OLS estimates (see Figure 6). We also observe that the dichotomous indicators create significantly smaller point estimates than the continuous indicator. As in Section 6.1, we argue that this difference can be explained with the low distinctiveness of the binary measures.

The 2SLS estimates differ in two notable ways from the OLS estimates (see Figure 7). First, we see that the size of the estimated effect of democracy on economic development increases if we set a higher threshold. The reason is a weaker first-stage relationship (see Appendix Table C.29). Second, the choice of the threshold value determines whether (and how) the scale of an index affects the regression result. If we use a low threshold value, the continuous indicator produces significantly larger 2SLS estimates than the dichotomous measure. For intermediate thresholds, we find no statistically significant difference. If the threshold is high, the estimates of the dichotomous index largely exceed the estimates of the continuous index.

Appendix Tables C.30 and C.31 show that we obtain similar results if we replace the multiplicative indicator with other indicators. The only minor change relates to the threshold at which the continuous and dichotomous index create similar 2SLS estimates since we observe that this threshold is not the same for all aggregation methods. Our results also hold when repeating the robustness checks that we proposed in Section 5.1.3 and Section 6.1.1 (see Appendix Tables C.32 – C.45).

In sum, 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" binary indicators should be used in regression analyses.[32]

7 Conclusion

In economics and other social sciences, building indices is the classical way of summarizing multidimensional data. A central challenge when designing social science indicators is to answer the question of how to transform the raw data into an unidimensional measure. The literature offers various data aggregation tools, but does not inform about the empirical consequences that arise from the decisions that researchers make when choosing a particular method. Using democracy as our expository example, we therefore investigate whether indices that differ with regard to their numerical form and aggregation function produce different results in regression analyses. Our results imply that both the choice between a continuous and dichotomous scale and the choice of the aggregation functions influences the results of OLS and 2SLS regressions in a statistically significant manner. The reason for why replacing one aggregation function by another has notable consequences is that the index values of the regimes at the upper and lower end of the autocracy-democracy spectrum crucially depend on the choice of the data aggregation tool. As a consequence, we find systematic differences in the extent to which indices change if a political transition takes place. Continuous and dichotomous indicators create different regression results because they differ in their discriminating power.

A question that might arise is whether the empirical consequences that we observe when comparing the performance of different democracy indicators also exist if we study another social science phenomena. For three reasons, we are convinced that this is indeed the case. First, the results of a simulation study that exploits randomly generated data to investigate the consequences of using

[32]Examples include the dichotomous Machine Learning index (see Section 6.1) and the index of Boix et al. (2013). The dichotomous index of Bjørnskov and Rode (2020) 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).

different aggregation methods are consistent with the results of our real data analysis. Second, the systematic differences between our continuous measure of democracy can be explained by the assumptions of the underlying aggregation functions. These assumptions do not change when we focus on another social science phenomena. Finally, the fact that continuous and dichotomous indices differ in their distinctiveness holds regardless of the considered phenomena.

Our paper presents a battery of regression results, suggesting that democracy promotes long-run economic growth. We also show that the magnitude of the estimated effect crucially depends on the choice of aggregation procedure. An obvious question is thus which of the methods produces the estimate that is closest to the true effect of democracy on economic development. Answering this question is difficult since we neither know the true effect nor the true level of democracy of all regimes. To address this problem, we present a simulation in which we focus on regimes for which the true degree of democratization can be derived from mild assumptions. The results of this simulation suggest that the Machine Learning approach outperforms the other aggregation methods. The basic reason for this result is that the Machine Learning technique is neither very likely to produce implausibly high indices for highly autocratic regimes nor to create implausibly low indices for highly democratic regimes. The example cases presented in Figures 2 and 3 illustrate this feature (see Section 3.4).[33] When we compare binary and continuous indices, our conclusion is that the binary indices produces less precise estimates due to its relatively low discriminating power.

Our results give rise to the following recommendations. First, when producing an index, scholars should check whether their aggregation function fits together with their conceptual assumptions. A mismatch exists, for example, if scholars combine an additive aggregation function and a concept that treats each aspect as a necessary condition. Second, scholars should provide some figures that illustrate how their indicator behaves, especially at the upper and lower end of the spectrum. Third, the appropriateness of an aggregation method should be justified based on example cases. Fourth, robustness tests are indispensable: the reader needs to know how the estimates react if the preferred aggregation tool is replaced by another method. This transparency is especially important when effect sizes form the basis for policy recommendations. Finally, dichotomizing a continuous index is problematic since the thresholds are arbitrary and influence regression results.

[33] For the Soviet Union, we observe that the additive and the item-response approach produce indices that are considerably larger than 0 (see Figure 2). Under the (mild) assumption that regimes without political competition are autocracies, these index values are implausibly high. Similarly, for Switzerland, we see that the five indices that are not produced by the Machine Learning approach indicate a non-negligible lack of democracy, even after the introduction of female suffrage in 1971. We argue that these measures are implausible since from 1971 onward there have been no notable restrictions on political competition, political participation, or the freedom of opinion in Switzerland.

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.

Aidt, T. S. and Eterovic, D. S. (2011). Political competition, electoral participation and public finance in 20th century latin america. *European Journal of Political Economy*, 27(1):181–200.

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.

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

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

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

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

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

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.

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(3):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. (2000). On the relationship between economic freedom and economic growth. *European Journal of Political Economy*, 16(2):215–241.

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.

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

Goertz, G. (2020). *Social Science Concepts and Measurement: New and Completely Revised Edition*. 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.

Gründler, K. and Krieger, T. (2021). Using machine learning for measuring democracy: A practitioners guide and a new updated dataset for 186 countries from 1919 to 2019. *European Journal of Political Economy*, 102047.

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.

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

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

Gwartney, J. and Lawson, R. (2003). The concept and measurement of economic freedom. *European Journal of Political Economy*, 19(3):405–430.

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.

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.

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

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.

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

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.

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.

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.

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.

Voigt, S. (2012). How to measure the rule of law. *Kyklos*, 65(2):262–284.

Wuttke, A., Schimpf, C., and Schoen, H. (2020). When the whole is greater than the sum of its parts: On the conceptualization and measurement of populist attitudes and other multidimensional constructs. *American Political Science Review*, 114(2):356–374.

For online publication

A Support Vector Machines

Support Vector Machines (SVM) is a frequently used Machine Learning method designed for pattern recognition. SVM aims at revealing an unknown functional relationship $\mathfrak{F}\colon \mathcal{X} \to \mathcal{Y}$ that links the 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 $\mathcal{S} = \{(\mathbf{x}_i, y_i) \,|\, i = 1, \ldots, n\}$:

$$\mathfrak{F}(\mathbf{x}_i) \overset{!}{=} y_i \qquad \forall\, i = 1, ..., n. \tag{27}$$

In contrast to conventional tools of statistical modeling (such as Ordinary Least Squares or Generalized Methods of Moments) Machine Learning tools do not need prior assumptions about the shape of the functional relationship of 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 methods.[34] SVM belongs to the former type because its application requires observations for learning the rule that maps the inputs onto the output (Steinwart and Christmann, 2008).[35]

The mathematical foundations of the 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 binary 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.[36]

A.1 Support Vector Classification

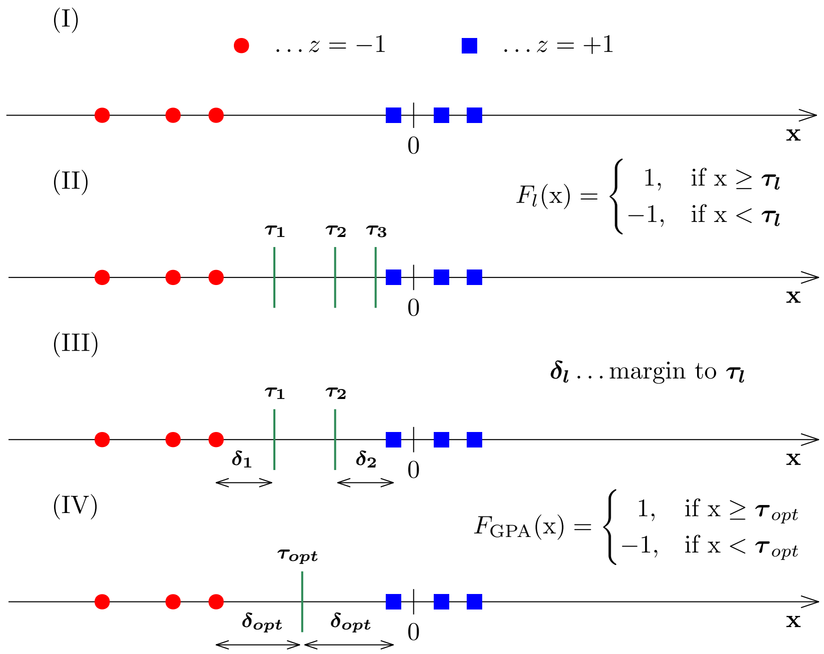
The Support Vector Classification (SVC) is a non- linear extension of the General Portrait Algorithm (GPA) developed by Vapnik and Lerner (1963) and Vapnik and Chervonenkis (1964). In its initial form, the GPA assumes the existence of some

[34]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 if no output variable exists and the available data need to be structured.

[35]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.

[36]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)

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



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

hyperplanes:

$$E_{\mathbf{w},b}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \qquad \mathbf{w} \in \mathbb{R}^m\,,\ ||\mathbf{w}|| = 1\,,\ b \in \mathbb{R}\,,\ \mathbf{x} \in \mathbb{R}^m \tag{28}$$

that can separate the observations in the sample $\mathcal{S}$ according to their labels $y \in \{-1, 1\}$.[37] Graph (I) in Figure A.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 A.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 computed. GPA selects the hyperplane with the greatest margin in $\mathcal{S}$ (Abe, 2005, Steinwart and Christmann, 2008). Graphs (III) and (IV) in Figure A.1 illustrate this procedure.

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

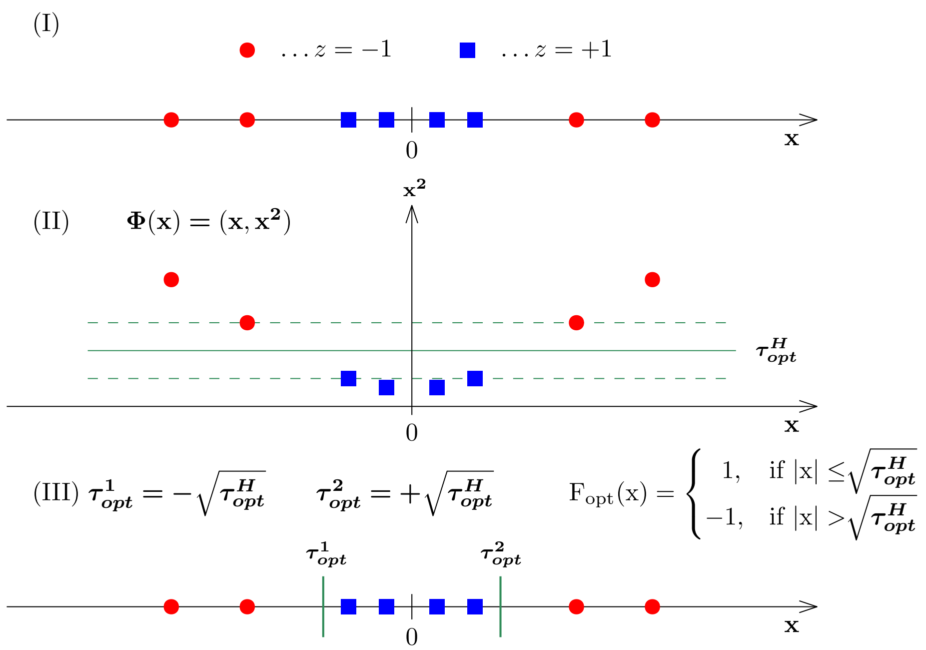
$$\min_{\mathbf{w},\,b}\ \frac{1}{2}\,\langle \mathbf{w},\, \mathbf{w} \rangle \qquad \text{s.t.} \qquad y_i \cdot (\langle \mathbf{w},\, \mathbf{x}_i \rangle\, +\, b)\, \geq\, 1 \tag{29}$$

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

$$\mathfrak{F}(\mathbf{x})\, =\, \mathrm{sign}\,(\langle \mathbf{w}^*, \mathbf{x} \rangle + b^*) \qquad \text{where} \qquad \mathbf{w}^* \,\in\, \mathbb{R}^m \quad \text{and} \quad b^* \,\in\, \mathbb{R}. \tag{30}$$

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

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



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

The GPA attracts little attention in applied research since a linear separation usually does not exist (see Graph (I) in Figure A.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* $\mathcal{H}$.[38] The GPA is then applied to the adjusted sample $\mathcal{S}_{\mathcal{H}} = \{(\Phi(\mathbf{x}_i), z_i) \,|\, i = 1, \dots, n\}$ and a dividing hyperplane is computed in $\mathcal{H}$:

$$E^{\mathcal{H}}_{\mathbf{w}^*_{\mathcal{H}}, b^*_{\mathcal{H}}}(\Phi(\mathbf{x})) = \langle \mathbf{w}^*_{\mathcal{H}}, \Phi(\mathbf{x}) \rangle + b^*_{\mathcal{H}} \quad \text{with} \quad \mathbf{w}^*_{\mathcal{H}} \in \mathcal{H} \quad \text{and} \quad b^*_{\mathcal{H}} \in \mathbb{R}. \tag{31}$$

The resulting classification function

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}\left(\langle \mathbf{w}^*_{\mathcal{H}}, \Phi(\mathbf{x}) \rangle + b^*_{\mathcal{H}}\right) \quad \text{with} \quad \mathbf{w}^*_{\mathcal{H}} \in \mathcal{H} \quad \text{and} \quad b^*_{\mathcal{H}} \in \mathbb{R} \tag{32}$$

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

Cortes and Vapnik (1995) argue 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

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

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

$$\min_{\mathbf{w}_{\mathcal{H}},\, b_{\mathcal{H}},\, \xi} \ \frac{1}{2}\langle \mathbf{w}_{\mathcal{H}},\, \mathbf{w}_{\mathcal{H}} \rangle \,+\, C \cdot \sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad y_i \cdot (\langle \mathbf{w}_{\mathcal{H}},\, \Phi(\mathbf{x}_i) \rangle \,+\, b_{\mathcal{H}}) \,\geq\, 1 \,-\, \xi_i \quad \forall\, i \tag{33}$$

where C denotes a fixed cost parameter for penalizing misclassifications.

If the dimension of $\mathcal{H}$ is large, solving problem (33) may turn out to be computationally infeasible (Steinwart and Christmann, 2008). In this case, the corresponding dual program:

$$\max_{\alpha \,\in\, [0,C]^n} \ \sum_{i=1}^{n} \alpha_i \,-\, \frac{1}{2} \cdot \sum_{i,j=1}^{n} z_i \cdot z_j \cdot \alpha_i \cdot \alpha_j \cdot \langle \Phi(\mathbf{x}_i),\, \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \quad \text{s.t.} \quad \sum_{i=1}^{n} z_i \cdot \alpha_i \,=\, 0 \tag{34}$$

can be used. In (34), $\alpha_1, \ldots, \alpha_n$ denote the Lagrange multipliers of the primal program. The dual program implies a closed form solution for the classification function:

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n} z_i \alpha_i^* \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}} \,+\, b_{\mathcal{H}}^* \right). \tag{35}$$

Because an appropriate *feature map* $\Phi\colon \mathcal{X} \to \mathcal{H}$ is usually unknown, Schölkopf et al. (1998) use 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* $\mathfrak{K}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$:

$$\mathfrak{F}(\mathbf{x}) \;=\; \operatorname{sign}\left(\sum_{i=1}^{n} z_i \cdot \alpha_i^* \cdot \mathfrak{K}(\mathbf{x}_i,\, \mathbf{x}) \,+\, b_{\mathcal{H}}^* \right). \tag{36}$$[39]

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 the Support Vectors affect the shape of the classification function (Abe, 2005, Steinwart and Christmann, 2008).

A.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 $\mathcal{S} = \{(\mathbf{x}_i,\, y_i) \,|\, i \,=\, 1, \ldots, n\}$:

$$|\, \mathfrak{F}(\mathbf{x}_i) \,-\, y_i \,| \overset{!}{\leq} \varepsilon \qquad \forall\, i = 1, \ldots, n. \tag{37}$$

[39] 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

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

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

$$\mathfrak{F}(\mathbf{x}) \;=\; \langle \mathbf{w},\, \mathbf{x} \rangle \,+\, b \qquad \text{with} \qquad \mathbf{w} \in \mathbb{R}^m \quad \text{and} \quad b \in \mathbb{R} \tag{38}$$

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

$$\min_{\mathbf{w},\, b} \; \frac{1}{2} \cdot ||\, w \,||^2 \qquad \text{s.t.} \qquad \begin{cases} z_i \,-\, \langle \mathbf{w},\, \mathbf{x}_i \rangle \,-\, b \,\leq\, \varepsilon \quad \forall\, i \\ \langle \mathbf{w},\, \mathbf{x}_i \rangle \,+\, b \,-\, z_i \,\leq\, \varepsilon \quad \forall\, i. \end{cases} \tag{39}$$

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

Since solving the constrained optimization problem (39) 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:

$$\min_{\mathbf{w},\, b,\, \xi_i^+,\, \xi_i^-} \; \frac{1}{2}\, ||\, w \,||^2 + C \cdot \sum_{i=1}^{n} \left(\xi_i^+ + \xi_i^-\right) \quad \text{s.t.} \quad \begin{cases} z_i \,-\, \langle \mathbf{w},\, \Phi(\mathbf{x}_i) \rangle \,-\, b \,\leq\, \varepsilon \,+\, \xi_i^+ \\ \langle \mathbf{w},\, \Phi(\mathbf{x}_i) \rangle \,+\, b \,-\, z_i \,\leq\, \varepsilon \,+\, \xi_i^- \\ \xi_i^+,\, \xi_i^- \,\geq\, 0. \end{cases} \tag{40}$$

To avoid computational issues if 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:

$$\mathfrak{F}(\mathbf{x}) \;=\; \sum_{i=1}^{n} \left(\alpha_i^+ \,-\, \alpha_i^-\right) \cdot \langle \Phi(\mathbf{x}_i),\, \Phi(\mathbf{x}) \rangle_{\mathcal{H}} \,+\, b_{\mathcal{H}}^*. \tag{41}$$

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

$$\mathfrak{F}(\mathbf{x}) \;=\; \sum_{i=1}^{n} (\alpha_i \,-\, \alpha_i^*) \cdot \mathfrak{K}(\mathbf{x}_i,\, \mathbf{x}) \,+\, b_{\mathcal{H}}^*. \tag{42}$$

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

B Background information for Section 5.3

In Section 5.3, we present a simulation analysis to examine which of our six aggregation tools performs best in regression analyses. To answer this question, we produce two sets of pseudo regimes. This supplementary section sketches the underlying data generation processes.

B.1 Autocratic pseudo regimes

The literature on the measurement of democracy widely agrees that a regime in which no political competition exists is non-democratic ($\Delta = 0$). Based on this consensus, we produce 1,000 autocratic pseudo regimes. Thereby, we proceed as follows:

- The absence of political competition implies that our four objective regime characteristic are equal to 0 and that our expert-based measure of party pluralism either indicates that there no political parties at all or that there is just one political party.
- Regarding the three regime characteristics that are related to political participation, we do not make any restrictions and choose them based on uniformly distributed random number generators. Our approach is consistent with the fact that political participation is extremely high in some non-competitive regimes (e.g. Cuba, Soviet Union) and absent in others (e.g. China, Saudi Arabia).
- With regard to our two regime characteristics on the freedom of opinion, we assume that the freedom of discussion for men and women is either not respect or weakly respected in non-competitive autocracies. The actual values for a particular autocratic pseudo regime are determined by uniformly distributed random number generators.

B.2 Democratic pseudo regimes

In our simulation analysis, we consider a regime as fully democratic ($\Delta = 1$) if there are no notable restrictions with regard to political competition, political participation, and the freedom of opinion. The ten regime characteristics of the 1,000 democratic pseudo regimes are thus computed as follows.

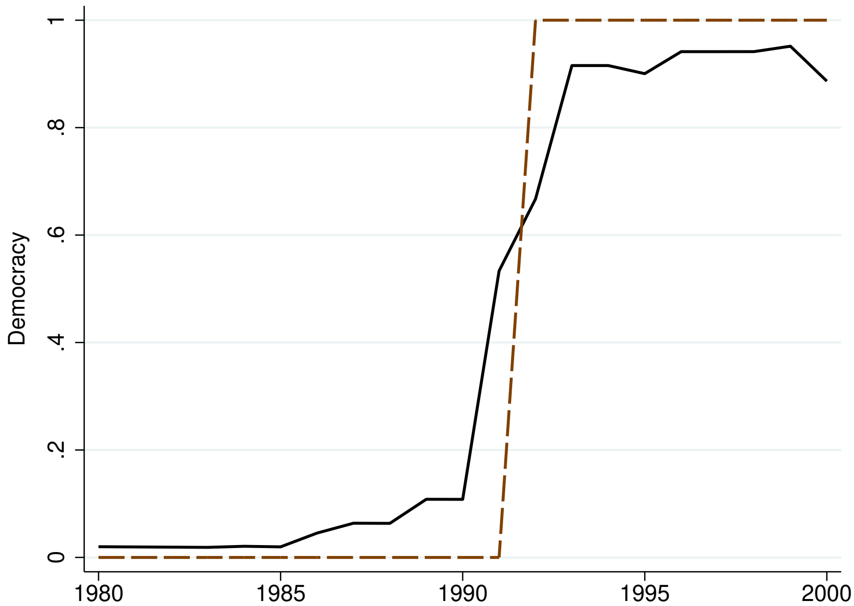
- The measure of party pluralism is set to the highest level, indicating a multiparty regime and the absence of any obstacles for opposition parties. The share of seats and the share of votes won by the strongest party is assumed to vary between 10 and 50 percent. The share of seats and the share of votes won by the runner-up party is at least half as large as the respective values of the leading party.[40] The actual values for our four objective regime characteristics on political competition are chosen via uniformly distributed random number generators.

- The suffrage in our democratic pseudo regimes is universal. When computing the turnout rate and the voter-population-ratio, we assume that at least 50 percent of the voting age population participate in an election. Conditional on this restriction, the actual values for these two regime characteristics are randomly drawn, using uniformly distributed random number generators.

- With regard to the freedom of discussion, we define that there are neither restrictions for men nor for women in our democratic pseudo regimes. The two corresponding regime characteristics are thus set to their highest level.

C Additional Tables and Figures

Figure C.1 Democracy in the Soviet Union and the Russian Federation (1980 – 2000).



Notes: The figures show the level of democracy of the Russian Federation and the Soviet Union, depending on whether we use the continuous Machine Learning index (solid black line) or the dichotomous Machine Learning index (dashed brown line). From 1991 onward, the measures of democracy refer to the Russian Federation.

[40] We admit that the thresholds are somehow arbitrary. During our analysis, we tried a variety of thresholds. The results presented in Section 5.3 hardly change if we modify the thresholds.

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)
Wald test (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)
Wald test (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 5) 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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 5) 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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 5) 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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 5) 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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 5) 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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 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 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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 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.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
Wald test (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 4) 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)
Wald test (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)
Wald test (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 5) is that we use the regime characteristics proposed by Teorell et al. (2019) in this robustness check.

Table C.17 Consequences of using different aggregation functions — OLS estimates (alternative weights).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	0.017^{***}	0.039^{***}	0.033^{***}	0.024^{***}	0.031^{***}	0.023^{***}
	(0.003)	(0.006)	(0.005)	(0.004)	(0.005)	(0.004)
Observations	10026	10026	10026	10026	10026	10026
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
Wald test (p-val.)	–	0.000	0.000	0.044	0.000	0.097

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 4) is that we weight all regime characteristics equally rather than using a PCA.

Table C.18 Consequences of using different aggregation functions — 2SLS estimates (alternative weights).

	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^{***}	0.063^{***}	0.067^{***}	0.040^{***}	0.049^{***}	0.039^{***}
	(0.006)	(0.011)	(0.012)	(0.009)	(0.009)	(0.007)
Wald test (p-val.)	–	0.000	0.000	0.044	0.000	0.097
	Panel B: First-stage estimates					
Demo. (reg.)	0.945^{***}	0.488^{***}	0.455^{***}	0.768^{***}	0.628^{***}	0.791^{***}
	(0.086)	(0.043)	(0.042)	(0.069)	(0.055)	(0.071)
Wald test (p-val.)	–	0.000	0.000	0.042	0.000	0.076
Observations	10026	10026	10026	10026	10026	10026
Countries	163	163	163	163	163	163
SW (F-stat.)	118.9	127.6	116.0	121.5	129.0	121.0
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	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. 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 4) is that we weight all regime characteristics equally rather than using a PCA.

Table C.19 Consequences of using different aggregation functions — OLS estimates (bivariate model).

	Machine Learning	Additive	Item-Response	Multi-plicative	Add./ Multi. (Average)	Add./ Multi. (CD function)
	(1)	(2)	(3)	(4)	(5)	(6)
Democracy	1.283^{***}	1.964^{***}	1.915^{***}	1.554^{***}	1.758^{***}	1.511^{***}
	(0.119)	(0.182)	(0.189)	(0.143)	(0.159)	(0.142)
Observations	11800	11800	11800	11800	11800	11800
Countries	163	163	163	163		163
R-Squared	0.212	0.238	0.211	0.215	0.230	0.210
Wald test (p-val.)	–	0.000	0.001	0.060	0.003	0.110

Notes: This table presents OLS estimates from a bivariate regression. The dependent variable is the log of GDP per capita. 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: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table C.20 Consequences of using different aggregation functions — 2SLS estimates (bivariate regressions).

	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	2.148***	3.019***	3.272***	2.598***	2.767***	2.572***
	(0.136)	(0.182)	(0.208)	(0.163)	(0.170)	(0.162)
Wald test (p-val.)	–	0.000	0.000	0.006	0.000	0.009
	Panel B: First-stage estimates					
Demo. (reg.)	1.092***	0.778***	0.717***	0.903***	0.848***	0.913***
	(0.0167)	(0.0288)	(0.0311)	(0.0351)	(0.0318)	(0.0354)
Wald test (p-val.)	–	0.000	0.000	0.000	0.000	0.000
Observations	11329	11329	11329	11329	11329	11329
Countries	163	163	163	163	163	163
SW (F-stat.)	660.0	726.4	528.6	663.7	710.5	665.0
AR (p-val.)	0.000	0.000	0.000	0.000	0.000	0.000

Notes: This table presents 2SLS estimates from a bivariate regression. The dependent variable is the log of GDP per capita. 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: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

Table C.21 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
Wald test (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 8) is that we use five-year averages rather than annual data.

Table C.22 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
Wald test (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 8) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.23 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
Wald test (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 8) is the dependent variable.

Table C.24 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
Wald test (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 8) is the dependent variable.

Table C.25 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
Wald test (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 8) 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.26 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
Wald test (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 8) 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.27 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
Wald test (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 8) is that we use the regime characteristics proposed by Teorell et al. (2019) in this robustness check.

Table C.28 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
Wald (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.29 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)
Wald (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)
Wald (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 5 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.30 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.31 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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 5 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.32 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30) is that we use five-year averages rather than annual data.

Table C.33 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) is that we use five-year averages rather than annual data.

Table C.34 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.35 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) is that we additionally control for population growth, armed conflict, and the rule of law.

Table C.36 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30) is that we change the outcome variable.

Table C.37 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) is that we change the outcome variable.

Table C.38 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30) is that we change the outcome variable.

Table C.39 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) is that we change the outcome variable.

Table C.40 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30) 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.41 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) 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.42 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30) 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.43 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) 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.44 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.28 and C.30 is that we use the regime characteristics proposed by Teorell et al. (2019).

Table C.45 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)
Wald (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)
Wald (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)
Wald (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)
Wald (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)
Wald (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.29 and C.31) is that we use the regime characteristics proposed by Teorell et al. (2019).