CESifo WORKING PAPERS

6930
2018

February 2018

Machine Learning Indices, Political Institutions, and Economic Development

Klaus Gründler, Tommy Krieger

CESifo

Impressum:

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

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

CESifo *Working Paper No. 6930*
Category 6: Fiscal Policy, Macroeconomics and Growth

Machine Learning Indices, Political Institutions, and Economic Development

Abstract

We present a new aggregation method - called *SVM algorithm* - and use this technique to produce novel measures of democracy (186 countries, 1960-2014). The method takes its name from a machine learning technique for pattern recognition and has three notable features: it makes functional assumptions unnecessary, it accounts for measurement uncertainty, and it creates continuous and dichotomous indices. We use the SVM indices to investigate the effect of democratic institutions on economic development, and find that democracies grow faster than autocracies. Furthermore, we illustrate how the estimation results are affected by conceptual and methodological changes in the measure of democracy. In particular, we show that instrumental variables cannot compensate for measurement errors produced by conventional aggregation methods, and explain why this failure leads to an overestimation of regression coefficients.

JEL-Codes: C260, C430, N400, O100, P160, P480.

Keywords: democracy, development, economic growth, estimation bias, indices, institutions, machine learning, support vector machines.

Klaus Gründler
University of Würzburg
Department of Economics
Sanderring 2
Germany – 97070 Würzburg
klaus.gruendler@uni-wuerzburg.de

Tommy Krieger
University of Konstanz
Department of Economics
Universitätsstraße 10
Germany – 78464 Konstanz
tommy.krieger@uni-konstanz.de

February 23, 2018
We are grateful for the constructive comments from Toke Aidt, Norbert Berthold, Roger Congleton, Martin Gassebner, Kai Gehring, Jerg Gutmann, Stephan Maurer, Niklas Potrafke, Guido Schwerdt, and Heinrich W. Ursprung. We would also like to thank seminar participants at the Universities of Cambridge and Konstanz, and conference participants at the annual meeting of the Public Choice Society in New Orleans, the annual meeting of the European Public Choice Society in Budapest, the Spring Meeting of Young Economists in Halle, the Silvaplana Workshop in Political Economy, the CESifo Political Economy Workshop in Dresden, and the annual meeting of the Verein für Socialpolitik in Vienna. Finally, we are highly appreciative of the financial support from the Graduate School of Decision Science (University of Konstanz).

1 Introduction

The view that authoritarian institutions are more conducive to economic development than liberal-democratic institutions has recently gained greater popularity among voters and politicians (Plattner, 2015, Wike et al., 2017). The Hungarian Prime Minister, Victor Orbán, for example, argued in his speech on July 26, 2014, that:

> "*The defining aspect of today's world can be articulated as a race to figure out a way of organizing communities and to find the state that is most capable of making a nation competitive. This is why, [...] a trending topic in thinking is understanding systems that are not Western, not liberal, not liberal democracies, maybe not even democracies, and yet making nations successful.*"

and attracted global attention by concluding that:

> "*What all this exactly means [...] we have to abandon liberal methods and principles of organizing a society, as well as the liberal way to look at the world.*".[1]

The question of how democracy affects economic growth is the subject of numerous studies. Economic theory suggests different mechanisms. Some theories argue that a democratic transition accelerates economic growth because democratic regimes increase public spending on human capital and improve the rule of law. Other theories predict a deceleration due to efficiency losses and higher taxation (Acemoglu, 2008, Besley and Coate, 1998, Persson and Tabellini, 1994, Saint-Paul and Verdier, 1993).

The empirical literature reflects this ambiguity. Doucouliagos and Ulubaşoğlu (2008) compile the results from more than 80 studies on the democracy-growth-nexus and observe that:

> "*15 percent of the estimates are negative and statistically significant, 21 percent of the estimates are negative and statistically insignificant, 37 percent of the estimates are positive and statistically insignificant, and 27 percent of the estimates are positive and statistically significant.*"

The results of more recent—and econometrically more sophisticated—studies are also diverse: Acemoglu et al. (2014), Gründler and Krieger (2016), Madsen et al. (2015), and Papaioannou and Siourounis (2008) conclude that democracy promotes economic growth, whereas Aisen and Veiga (2013), Madsen and Murtin (2017), and Murtin and Wacziarg (2014) find no empirical support for this view.[2]

[1] For an English version of the speech, see `https://hungarianspectrum.wordpress.com/2014/07/31/`.

[2] A related strand of literature investigates the effect of economic development on democracy (Acemoglu et al., 2008, 2009, Barro, 1999, Cervellati et al., 2014, Gundlach and Paldam, 2009, Lipset, 1959, Przeworski, 2000) and the effect of income shocks on institutional change (Aidt and Franck, 2015, Aidt and Leon, 2016, Brückner and Ciccone, 2011, Brückner et al., 2012, Burke and Leigh, 2010, Chaney, 2013, Franck, 2016). This study does not examine these relationships. Our methodological contributions may, however, also be relevant for these lines of research.

Our study makes three contributions. First, we present new measures of democracy for 186 countries and the time period from 1960 to 2014.[3] To transform the employed raw data on political competition, electoral participation, and political discourse into a single index, we propose an aggregation method that is based on a machine learning technique for pattern recognition, known as Support Vector Machines (SVM). We argue that this methodological innovation is necessary since conventional aggregation techniques—such as arithmetic and geometric averaging, or latent variable methods—produce misleading indices for specific autocracies and established democracies. We also show that these systematic malfunctions result from functional assumptions about the relationship between the variables in the raw data set and the degree of democratization. The SVM algorithm avoids these assumptions and thus generates more reliable measures. Our method has two additional features: it computes continuous and dichotomous indices for any concept of democracy, and it provides indication of measurement uncertainty.[4]

Second, we provide new insights into the relationship between democracy and economic development. We address endogeneity problems with a common two-stage least squares (2SLS) approach in which the average degree of democratization in neighboring countries serves as the instrument for the domestic degree of democratization. The motivation for this instrumentation strategy is the observation that institutional transitions often occur in regional waves, and thus a correlation exists between the domestic and the regional level of democracy (Acemoglu et al., 2014). Our regression results show that a transition from an authoritarian regime towards a democratic regime leads to a statistically significant increase in economic development. This effect persists when controlling for other explanatory factors, applying alternative identification strategies, and taking account of measurement uncertainty in the SVM indices.[5]

Finally, we investigate the empirical consequences of changes in the measure of democracy. In this respect, we provide novel contributions along three dimensions. First, we compare continuous and dichotomous indices that are conceptually equivalent. We find that switching from one type to another can cause significant changes in the regression re-

[3] Measuring democracy is difficult for multiple reasons. In this study, we use a common three-stage procedure (Munck and Verkuilen, 2002): first, we define the term "democracy' (*conceptualization*), then compile data that reflect the components of the concept (*operationalization*), and finally specify the rule that transforms the raw data into an index (*aggregation*).

[4] A preliminary version of the SVM algorithm is used by Gründler and Krieger (2016) to synthesize existing measures of democracy. Our study features a number of methodological and conceptual improvements. We (i) use non-aggregated data and thus achieve greater precision with regard to conceptual issues, (ii) extend the length of the panel by 25 years, (iii) compute continuous and dichotomous indices, (iv) provide an indication of measurement uncertainty, and (v) increase the numerical robustness of the algorithm. This paper also provides a clear explanation for why the SVM algorithm outperforms conventional aggregation techniques.

[5] Existing empirical studies suggest that there are two main reasons for why democratic institutions are growth-enhancing: (i) democracies spend more on education and health services, and thus improve the accumulation of human capital (Ansell, 2010, Besley and Kudamatsu, 2006, Cascio and Washington, 2013, Fujiwara, 2015, Harding and Stasavage, 2013, Kudamatsu, 2012, Tavares and Wacziarg, 2001), and (ii) democratic regimes have economic institutions that are conducive to economic development (Besley et al., 2010, De Haan and Sturm, 2003, Giuliano et al., 2013, Grosjean and Senik, 2011, Lundström, 2005, Rode and Gwartney, 2012). The identification of the exact channel running from democracy to economic development is, however, beyond the scope of this paper.

sults. We also show that the estimated effect of democracy on economic development is less stable when using a dichotomous measure. Second, we modify the concept of democracy and observe that conceptual aspects are of minor importance for the estimation results. Third, we replace the SVM algorithm with conventional aggregation techniques and find notable increases in the estimated effect of democracy on economic development. We demonstrate that these increases can be explained by the measurement errors produced by conventional aggregation methods. We also explain why the 2SLS estimator does not prevent the overestimation.

The remainder is organized as follows: Section 2 provides an introduction to SVM. Section 3 derives the measures of democracy. Section 4 reports the empirical results. Section 5 concludes.

2 Support Vector Machines

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

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

In contrast to conventional tools of statistical modeling—such as Ordinary Least Squares (OLS) or Generalized Methods of Moments (GMM)—machine learning techniques do not require prior assumptions about the shape of the functional relationship, they rather learn without being explicitly programmed (Breiman et al., 2001). The literature distinguishes between supervised and unsupervised machine learning techniques. SVM belongs to the former type because its application requires a set of observations for learning the rule that maps $\mathbf{x}$ onto z (Steinwart and Christmann, 2008).[6]

The mathematical foundations of SVM techniques and their properties with regard to prediction accuracy, statistical robustness, practicability are well documented (Abe, 2005, Bennett and Campbell, 2000, Steinwart and Christmann, 2008). In this study, we employ two standard SVM techniques to arrive at dichotomous classifications and to run non-linear regressions. In the remainder of this section, we introduce the mathematical formulations of Support Vector Classification and Support Vector Regression.[7]

[6]In this context, "learning the rule" means that an empirical model is estimated which adequately predicts the output z 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 component x_j ($j = 1, \ldots, m$) on the outcome z (Steinwart and Christmann, 2008).

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

2.1 Support Vector Classification

A Support Vector Classification (SVC) is a non-linear extension of the General Portrait Algorithm (GPA). In its initial form, the GPA assumes the existence of hyperplanes

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

that separate the observations in $\mathcal{S} = \{(\mathbf{x}_i, z_i) \,|\, i = 1, \ldots, n\}$ according to their labels $z \in \{-1, 1\}$. Graph (I) in Appendix Figure B.1 illustrates this separation in a one-dimensional example.

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

In formal terms, the GPA solves the quadratic optimization problem

$$\min_{\mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R}} \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle \qquad \text{s.t.} \quad y_i \left(\langle \mathbf{w}, \mathbf{x}_i \rangle + b \right) \geq 1 \tag{3}$$

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

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

GPA attracts little attention in applied research because a linear separation usually does not exist (see Graph (I) in Appendix Figure B.2). Boser et al. (1992) extend the GPA to allow for the estimation of non-linear classification functions. They propose the usage 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* $\mathcal{H}$.[8] The GPA is then applied to the adjusted sample $\mathcal{S}_{\mathcal{H}} = \{(\Phi(\mathbf{x}_i), z_i) \,|\, i = 1, \ldots, n\}$ and a dividing hyperplane is computed in $\mathcal{H}$:

$$H^{\mathcal{H}}_{\mathbf{w}^*_{\mathcal{H}}, b^*_{\mathcal{H}}}(\Phi(\mathbf{x})) = \langle \mathbf{w}^*_{\mathcal{H}}, \Phi(\mathbf{x}) \rangle + b^*_{\mathcal{H}} \qquad \mathbf{w}^*_{\mathcal{H}} \in \mathcal{H} \,,\ b^*_{\mathcal{H}} \in \mathbb{R}. \tag{5}$$

The resultant classification function

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign} \left(\langle \mathbf{w}^*_{\mathcal{H}}, \Phi(\mathbf{x}) \rangle + b^*_{\mathcal{H}} \right) \qquad \mathbf{w}^*_{\mathcal{H}} \in \mathbb{R}^m \,,\ b^*_{\mathcal{H}} \in \mathbb{R} \tag{6}$$

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

Cortes and Vapnik (1995) argue that random noise and measurement errors may lead to

[8] 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).

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

$$\min_{\mathbf{w}_\mathcal{H}\in\mathcal{H}, b_\mathcal{H}\in\mathbb{R}, \xi\in\mathbb{R}^n_+} \frac{1}{2}\langle \mathbf{w}_\mathcal{H}, \mathbf{w}_\mathcal{H}\rangle + C\sum_{i=1}^{n}\xi_i \quad \text{s.t.} \quad z_i\left(\langle \mathbf{w}_\mathcal{H}, \Phi(\mathbf{x}_i)\rangle + b_\mathcal{H}\right) \geq 1 - \xi_i \;\forall\, i, \tag{7}$$

where C denotes a fixed cost parameter for penalizing misclassifications.

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

$$\max_{\alpha\in[0,C]^n} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} z_i z_j \alpha_i \alpha_j \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle_\mathcal{H} \quad \text{s.t.} \quad \sum_{i=1}^{n} z_i\alpha_i = 0 \tag{8}$$

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

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n} z_i\alpha_i^* \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x})\rangle_\mathcal{H} \; + \; b_\mathcal{H}^*\right). \tag{9}$$

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

$$\mathfrak{F}(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{n} z_i\alpha_i^* \mathfrak{K}(\mathbf{x}_i, \mathbf{x}) \; + \; b_\mathcal{H}^*\right).^{9} \tag{10}$$

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

2.2 Support Vector Regression

In their traditional form, GPA and SVC are limited to applications in which the output variable comes from a countably finite set. Vapnik (1995, 1998) overcomes this constraint by introducing a method that estimates real-valued functions. The objective of Support Vector Regression (SVR) is to find a function $\mathfrak{F}\colon \mathcal{X} \subseteq \mathbb{R}^m \to \mathcal{Z} \subseteq \mathbb{R}$ whose predicted outcomes deviate at most by ε from the observed labels for all observations in $\mathcal{S} = \{(\mathbf{x}_i, z_i) \,|\, i = 1, \ldots, n\}$:

$$|\mathfrak{F}(\mathbf{x}_i) - z_i| \overset{!}{\leq} \varepsilon \qquad \forall i = 1, \ldots, n. \tag{11}$$

[9] This 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

$$\mathfrak{F}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \qquad \mathbf{w} \in \mathbb{R}^m\,,\, b \in \mathbb{R}\,,\, \mathbf{x} \in \mathbb{R}^m \tag{12}$$

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

$$\min_{\mathbf{w}\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}||w||^2 \quad \text{s.t.} \quad \begin{cases} z_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon & \forall i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - z_i \leq \varepsilon & \forall i. \end{cases} \tag{13}$$

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

Since solving this constrained optimization problem often turns out to be impossible, the applicability of a linear SVR is limited. Vapnik (1995, 1998) therefore 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

$$\min_{\mathbf{w}\in\mathcal{H}, b\in\mathbb{R}, (\xi_i^+,\xi_i^-)\in\mathbb{R}_+^2} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\left(\xi_i^+ + \xi_i^-\right) \text{ s.t. } \begin{cases} z_i - \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^+ \\ \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b - z_i \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0. \end{cases} \tag{14}$$

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

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

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

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

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n}\left(\alpha_i^+ - \alpha_i^-\right)\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x})\rangle_{\mathcal{H}} + b_{\mathcal{H}}^*. \tag{15}$$

Since $\Phi\colon \mathcal{X} \to \mathcal{H}$ 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

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\,\mathfrak{K}(\mathbf{x}_i, \mathbf{x}) + b_{\mathcal{H}}^*. \tag{16}$$

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

3 Measuring democracy

We use a three-stage procedure to measure the degree of democratization (Munck and Verkuilen, 2002): first, we define the term "democratic regime" (*conceptualization*), then compile data that reflect the components of the concept (*operationalization*), and finally specify the rule that transforms the raw data into an index (*aggregation*). In formal terms, this simply means that the degree of democratization is a function of the observed characteristics, $\mathbf{x} = (x_1, \ldots, x_m) \in \mathcal{X}$:

$$d^{\mathfrak{D}} := \mathfrak{F}^{\mathfrak{D}}(\mathbf{x}) = \mathfrak{F}^{\mathfrak{D}}(x_1, \ldots, x_m) \in \mathcal{D} \subseteq [0,1] \qquad \forall \mathbf{x} \in \mathcal{X} \subseteq [0,1]^m, \tag{17}$$

where $\mathfrak{F}^{\mathfrak{D}} \colon \mathcal{X} \to \mathcal{D}$ denotes the aggregation function for a given concept $\mathfrak{D}$.

The implementation of Equation (17) entails two methodological challenges: First, we need to ascertain whether the intended application of the index requires a continuous or a dichotomous codomain $\mathcal{D}$. Second, and even more challenging, we need to specify the function that aggregates the characteristics. This task is difficult because the correct aggregation rule is not known and the actual degrees of democratization cannot be observed directly.

The conventional procedure for solving the aggregation problem is, of course, to make assumptions about the functional relationship between the characteristics and the degree of democratization. Scholars have usually first assigned a weight $\omega_j \in (0,1)$ to each characteristic x_j $(j = 1, \ldots, m)$, and then chosen a functional link that is either additive or multiplicative (Decancq and Lugo, 2013, Goertz, 2006):

$$\mathfrak{F}^{\mathfrak{D}}(\mathbf{x}) = \sum_{j=1}^{m} \omega_i \cdot x_i \quad \text{or} \quad \mathfrak{F}^{\mathfrak{D}}(\mathbf{x}) = \prod_{j=1}^{m} x_j^{\omega_j} \quad \text{with} \quad \sum_{j=1}^{m} \omega_j = 1.^{10} \tag{18}$$

This conventional approach has three major weaknesses:[11] the weighting schemes and functional forms can hardly be justified with the help of the theory of democracy (Munck and Verkuilen, 2002), measurement uncertainty that arises from data limitations cannot be indicated since conventional aggregation methods do not allow for the calculation of confidence intervals (Pemstein et al., 2010), and the additive and multiplicative linking leads to distortions at the lower and at the upper ends of the spectrum (Teorell et al., 2016).[12]

3.1 Measuring democracy with SVM

This section introduces a new procedure for measuring democracy. The SVM algorithm has three major advantages compared to conventional aggregation techniques: (i) assump-

[10] Some more recent studies suggest synthesizing additive and multiplicative indices (Acemoglu et al., 2014, Pemstein et al., 2010, Teorell et al., 2016). For details, see Section 3.5.

[11] Appendix A.1 illustrates the limitations of the conventional procedure using the democracy indicators developed by Acemoglu et al. (2014), Marshall et al. (2016) and Vanhanen (2000).

[12] Sections 3.5 and 4.4 show why these distortions arise and why they are problematic.

tions about the functional relationship between the characteristics and the degree of democratization are not necessary, (ii) continuous and dichotomous measures can be created without manual adjustments and for any concept of democracy, and (iii) a distribution of indicators can be produced to provide indication of measurement uncertainty.

Conceptualization and operationalization

In the ***first step*** of the procedure, we select institutional components to define the term "democratic regime". The selection of suitable components is much debated issue and generally acknowledged rules are certainly not available (Munck and Verkuilen, 2002). We therefore propose a procedure that is applicable to all concepts of democracy, i.e. the method adjusts the aggregation function to fit the needs of the chosen definition. In this respect, we follow Guttman (1994) who suggest that there is no point in arguing about what the correct definition is. The baseline concept is limited to the procedural aspects of democracy and is thus based on the components "electoral participation", "political competition", and "political discourse" (Collier and Levitsky, 1997).[13]

In the ***second step***, we operationalize our baseline concept of democracy by selecting suitable characteristics. We follow the recommendations by Munck and Verkuilen (2002), who suggest the use of non-aggregated data from subjective and objective sources, and restrict ourselves to characteristics which are available for a comprehensive sample of country-years.

We define electoral participation as the right of citizens to elect their political leaders (Dahl, 1971). Suffrage may be limited, either through constitutional restrictions that exclude citizens because of their gender, race, or social status, or by non-constitutional restrictions resulting from civil war, martial law, and political repression. We collect data on voter turnout and the voter-population-ratio to capture both types of disenfranchisement.[14]

We relate political competition to meaningful contests among individuals and parties

[13] The advantages and disadvantages of different concepts are well documented (Collier and Levitsky, 1997, Munck and Verkuilen, 2002, O'Donnell, 2001). We refrain from taking a position in the debate on the correct definition of democracy. We simply use a standard concept to illustrate how SVM can be applied for the computation of indices. Section 4.4 shows how a modification of the concept affects the output of the SVM algorithm.

[14] Voter turnout is defined as the ratio of active voters to registered voters. The number of voters is 0 when the government or parliament is not legitimized by a general election. We obtain the data from different sources, including: African Election Database (2017), Carr (2017), IDEA (2017), IFES (2017), IPU (2017), Nohlen (2005), Nohlen et al. (2001, 1999), and Nohlen and Stöver (2010). A documentation of the collected data is available upon request.

We are aware that voter turnout is often criticized as a weak measure of electoral participation (Munck and Verkuilen, 2002). The literature highlights two problems: First, voter turnout does not identify the existence of constitutional disenfranchisement (Bollen, 1980). For instance, voter turnout in South Africa rarely falls below 50 percent, even though the Apartheid regime excluded more than 70 percent of adult citizens from the election process (Nohlen et al., 1999). We address this problem by using the voter-population-ratio as an additional characteristic of electoral participation. Second, Hadenius (1992) argues that a high turnout rate does not necessarily correlate with democratic institutions. He supports his argument with the example of the Soviet Union, where voter turnout always exceeded 99 percent. We share this concern, but emphasize that it does not apply to the SVM algorithm since our method allows for a non-linear relationship between the turnout rate and the degree of democratization.

running for public office (Diamond et al., 1990). The operationalization relies on objective and subjective data and includes five characteristics: (i) an expert-based rating on party pluralism[15], (ii) the share of votes that are not won by the leading party, (iii) the share of parliamentary seats that are not won by the leading party, (iv) the number of votes won by the leading opposition party divided by the number of votes won by the leading party, and (v) the number of parliamentary seats occupied by the leading opposition party divided by the number of parliamentary seats occupied by the leading party.[16]

We apply expert-based information on the freedom of discussion to operationalize the quality of political discourse. The ratings are gender-specific and indicate on a five-point scale the extent to which men (women) are able to debate political issues in public spaces and private homes. The data was collected by Coppedge et al. (2016) and is currently the gold standard in political science.

Priming data

To use SVM for the measurement of democracy, we need observations—called *priming data*—with a generally acknowledged degree of democratization. In the ***third step***, we therefore select country-years that are suited for this purpose. We proceed from the following presumptions: (i) the most and least democratic regimes are uncontroversial and thus suitable,[17] and (ii) a country-year belongs to the group of the least (most) democratic regimes when it is in the lower (upper) decile of the index developed by Pemstein et al. (2010), or when it is in the lower (upper) decile of the index developed by Teorell et al. (2016).[18]

---

[15] The index of party pluralism is ordinal and has five categories: (i) there are no political parties, (ii) there is one political party, (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 (iv) there are multiple parties and virtually no obstacles for opposition parties. The core data come from Coppedge et al. (2016) and are supplemented by information from African Election Database (2017), IPU (2017), Nohlen (2005), Nohlen et al. (2001, 1999), and Nohlen and Stöver (2010).

[16] If the government or parliament is not legitimized by a general election, the four objective characteristics are equal to 0. We obtain the objective data from different sources, including: African Election Database (2017), Carr (2017), IFES (2017), IPU (2017), Nohlen (2005), Nohlen et al. (2001, 1999), and Nohlen and Stöver (2010).

We use separate measures for votes and parliamentary seats because they relate to different aspects of political competition. The former reflects information on public competition, while the latter focuses exclusively on parliamentary competition. The two types of competition can vary substantially from one another. The Mongolian People's Revolutionary Party, for example, won about 51.6 percent of the votes and 94.7 percent of the parliament seats in the general election in 2002 (Nohlen et al., 2001). We use information on the leading opposition party in order to distinguish between countries in which the ruling party is faced with fragmented opposition and countries in which the leading party is faced with a serious political opponent.

[17] According to Cheibub et al. (2010) and Lindberg et al. (2014), distinguishing the least democratic from the most democratic regimes is a rather simple exercise: most scholars agree that Sweden is a democracy and Saudi Arabia is not. We exploit the consensus on regimes types at the ends of the spectrum to justify the assumption that their degree of democratization is directly observable.

[18] We use the measures of democracy developed by Pemstein et al. (2010) and Teorell et al. (2016) because of their scaling and availability. Section 3.3 addresses potential concerns about our selection strategy. Appendix Table C.1 lists the 2415 country-years that are included in the priming data.

Aggregation

In the ***fourth step***, we decide upon the codomain of the measure of democracy. We can choose between a continuous scale ($\mathcal{D} = [0,1]$) and a dichotomous scale ($\mathcal{D} = \{0,1\}$).[19] In the ***fifth step***, we randomly select country-years from the priming data to create the training set $\mathcal{T}_\zeta$. In the ***sixth step***, we use SVM to estimate either the classification function or the regression function:

$$\mathfrak{F}_{\mathcal{T}_\zeta}^{\mathfrak{D}\{0,1\}} : \mathcal{X} \to \{0,1\} \quad \text{or} \quad \mathfrak{F}_{\mathcal{T}_\zeta}^{\mathfrak{D}[0,1]} : \mathcal{X} \to [0,1]\,.^{20}$$

In the ***seventh step***, we use the classification or regression function to compute a dichotomous or continuous index for each country-year in the data set

$$d_{it_\zeta}^{\mathfrak{D}\{0,1\}} = \mathfrak{F}_{\mathcal{T}_\zeta}^{\mathfrak{D}\{0,1\}}(\mathbf{x}_{it}) \quad \text{or} \quad d_{it_\zeta}^{\mathfrak{D}[0,1]} = \mathfrak{F}_{\mathcal{T}_\zeta}^{\mathfrak{D}[0,1]}(\mathbf{x}_{it}) \qquad \forall(i,t).$$

In the ***final step*** of the algorithm, we repeat the process from Step 5 to Step 7 for all iterations $\zeta \in \{0,\ldots,\zeta_{max}\}$.[21]

For each country-year, we can extract two pieces of information from the SVM algorithm: (i) the *Dichotomous* or *Continuous Support Vector Machines Democracy Index (DSVMDI, CSVMDI)*, i.e. the median of all realizations

$$d_{it}^{\mathfrak{D}\{0,1\}} = \text{med}_{\zeta=0,\ldots,\zeta_{max}}\left[d_{it_\zeta}^{\mathfrak{D}\{0,1\}}\right] \quad \text{or} \quad d_{it}^{\mathfrak{D}[0,1]} = \text{med}_{\zeta=0,\ldots,\zeta_{max}}\left[d_{it_\zeta}^{\mathfrak{D}[0,1]}\right] \quad \forall(i,t),$$

and (ii) the *Dichotomous* or *Continuous SVM Distribution (DDist, CDist)*, which represent the sets of all percentiles.

3.2 Measurement uncertainty

Since data for the characteristics $\mathbf{x} = (x_1,\ldots,x_m)$ is limited and prone to error, there is a trade-off between the quality of the information, the sample size, and the number of characteristics. We give high priority to data quality and follow Munck and Verkuilen (2002), who recommend using non-aggregated data from both subjective and objective sources (see Section 3.1). However, despite our best efforts, inaccuracies remain a problem for some observations due to interpolated and inconsistent data.

[19] Social scientists disagree about whether a continuous or dichotomous measure is more capable of investigating the relationship between democracy and economic development. We have no preconceived opinion on this issue and refrain from taking sides. Section 4.4 compares the empirical properties of the two scaling types.

[20] To use SVM techniques, we have to specify the penalization parameter C and the kernel function (see Section 2). We set $C = 1$ as proposed by Mattera and Haykin (1999) and use the Gaussian RBF kernel function since it provides the most promising results in our robustness checks. To perform a Support Vector Regression, we also need to calibrate the margin parameter ε (see Section 2.2). For this purpose, we follow Ragin (2008), who sets the degree of democratization of highly autocratic (democratic) regimes equal to 0.05 (0.95). We use these figures as benchmarks for the country-years in the priming data and set $\varepsilon = 0.05$. This choice implies that the level of democracy of autocratically (democratically) labeled observations can vary between 0 and 0.1 (0.9 and 1) without any penalization.

[21] The number of iterations is set to $\zeta_{max} = 2000$. The convergence of the SVM algorithm is ensured through this choice.

Table 1 Measurement uncertainty — Imprecise raw data

	Continuous SVM Indices					Dichotomous SVM Indices				
Distortion (λ)	0.01	0.025	0.05	0.1	0.2	0.01	0.025	0.05	0.1	0.2
Mean Deviation	0.001	0.002	0.003	0.006	0.011	0.001	0.001	0.003	0.005	0.010
Median Deviation	0.001	0.001	0.002	0.004	0.007	0.000	0.000	0.000	0.000	0.000
Correlation	1.000	1.000	1.000	1.000	0.999	0.999	0.997	0.993	0.989	0.979
Coverage Rate	1.000	1.000	0.998	0.991	0.946	1.000	1.000	1.000	1.000	0.999

Notes: The SVM indices suffer from measurement uncertainties because of interpolated and inconsistent data on the characteristics. This table shows that the distributions produced by the SVM algorithm (CDist, DDist) reflect these uncertainties. To obtain these results, we first generate randomly biased characteristics (for details, see Equation 19). The parameter λ specifies the degree of distortion: for example, $\lambda = 0.1$ implies that voter turnout can be biased up to ± 3.4 percent. Next, we apply the SVM algorithm to the biased characteristics and thus compute randomly biased SVM indices. We report four statistics to summarize the comparisons between the baseline indicators ($\lambda = 0$) and the biased indicators ($\lambda > 0$): (i) the mean of the absolute deviations, (ii) the median of the absolute deviations, (iii) the correlation coefficient, and (iv) the coverage rate, i.e. the share of country-years for which the biased indicator is between the upper and lower bounds of the baseline distribution.

We argue that the estimated distributions (CDist, DDist) indicate the measurement uncertainty that results from these data problems. To substantiate this hypothesis, we proceed in two steps: First, we generate randomly biased characteristics

$$\tilde{x}_j \;=\; x_j \;+\; \nu_j \quad \text{with} \quad \nu_j \;\sim\; \mathcal{U}(-\lambda \cdot \sigma_{\mathbf{x}_j}, +\lambda \cdot \sigma_{\mathbf{x}_j}), \tag{19}$$

where λ is a parameter to specify the degree of distortion and $\sigma_{\mathbf{x}_j}$ denotes the sample standard deviation of characteristic x_j ($j = 1, \ldots, m$). Second, we apply the SVM algorithm and compare the biased SVM indices ($\lambda > 0$) with the baseline measures of democracy ($\lambda = 0$).

Table 1 presents the results of the two-step analysis. We report four summary statistics: (i) the correlation between the baseline and the biased SVM indices, (ii) the average of the absolute deviations, (iii) the median of the absolute deviations, and (iv) the coverage rate, i.e. the share of country-years for which the biased SVM index is captured by the baseline distribution. Two findings are particularly relevant: First, the SVM distributions largely absorb the measurement uncertainties that originate from imprecise raw data. Second, since deviations are, on average, smaller and coverage rates are larger for DSVMDI, the SVM algorithm supports the hypothesis advanced by Alvarez et al. (1996), who suspect that continuous indicators are less stable than dichotomous indicators.

3.3 Selection of the priming data

3.3.1 Adequacy of the labeling

The priming data include 2415 country-years: 1161 democracies and 1254 autocracies (for a complete list, see Appendix Table C.1). We use the election handbooks prepared by Dieter Nohlen (1999, 2001, 2005, 2010) and historical records to check whether the assigned labels are consistent with our concept of democracy. We detected only two observations for which the classification may be controversial: The first is the observation for Switzerland in 1970, because women were not entitled to vote in national elections. The second is the observation for Israel in 1999, since Palestinians living in the Gaza Strip and the West

Table 2 Priming data — Agreement rates

	ANRR		BMR		CGV		SGB		Polity		Polyarchy	
	auto	demo	auto	demo	auto	demo	auto	demo	auto	demo	auto	demo
Label (auto)	1.000	0.000	1.000	0.000	1.000	0.000	0.958	0.000	0.983	0.000	0.992	0.000
Label (demo)	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.999	0.000	1.000	0.000	0.999

Notes: To use SVM for the measurement of democracy, we need observations—called *priming data*—with a generally acknowledged degree of democratization. To select country-years that are suited for this purpose, we proceed from two presumptions: (i) the most and least democratic regimes are uncontroversial and thus suitable, and (ii) a country-year belongs to the group of the least (most) democratic regimes when it is in the lower (upper) decile of the index developed by Pemstein et al. (2010), or when it is in the lower (upper) decile of the index developed by Teorell et al. (2016). This table substantiates these assumptions, using six alternative measures of democracy. We report agreement rates, i.e. the share of selected country-years that are classified as *autocratic* or *democratic* by an alternative indicator. The list of alternative measures of democracy includes the dichotomous indicators developed by Acemoglu et al. (2014) [ANRR], Boix et al. (2013) [BMR], and Cheibub et al. (2010) [CGV], and the continuous indicators developed by Skaaning et al. (2015) [SGB], Marshall et al. (2016) [Polity], and Vanhanen (2000) [Polyarchy]. A regime is classified as *autocratic* (*democratic*) if Polity < -5 (> 5); if Polyarchy < 5 (> 15); or if SGB < 2 (> 5).

Bank were not allowed to participate in the general election for the Israeli parliament.[22]

Furthermore, we use six alternative measures of democracy to investigate whether a consensus exists with regard to the labeled observations. Table 2 reports the agreement rates and confirms that there is little reason to doubt the adequacy of the labeling.

3.3.2 Representativeness of the priming data

Autocratic and democratic regimes can take different forms (Cheibub et al., 2010). The priming data should reflect this institutional heterogeneity, since otherwise we may produce SVM indices that discriminate in favor of specific types of government. In the following, we demonstrate that this basic prerequisite is satisfied.

Cheibub et al. (2010) argue that autocracies can be classified according to the sociopolitical background of the rulers. Their database allows us to differentiate between four types of authoritarianism: civil dictatorship, communist dictatorship, military dictatorship, and royal dictatorship. Column 1 of Table 3 shows that each of these forms of autocratic governance is represented in the priming data.

Schedler (2002) discerns between electoral autocracies and closed autocracies: while the former category refers to autocracies in which non-competitive elections take place, the latter category encompasses all regimes in which the government does not organize national elections. Columns 2 and 3 of Table 3 indicate that the priming data includes autocracies of both types.

The theory of dictatorship presented by Olson (1993) implies that durability can be used to classify autocratic regimes. We distinguish between three levels of durability: (i) autocracies that exist for less than five years, (ii) autocratic regimes with a duration of between 6 and 25 years, and (iii) dictatorships that are in power for more than 25 years. Columns 4 to 6 of Table 3 illustrate that the autocratically labeled observations are distributed across all categories.

Cheibub et al. (2010) distinguishes between three forms of democratic governance: (i)

[22]Note that the SVM indices remain virtually unchanged when we exclude these two country-years from the priming data.

Table 3 Priming data — Institutional heterogeneity among autocratic regimes

	Regime	**Election**		**Duration in Years**		
	Σ	No	Yes	≤ 5	≤ 25	> 25
Civil dictatorship	260	170	90	83	159	18
Communist dictatorship	300	149	151	16	101	183
Military dictatorship	424	312	112	161	222	41
Royal dictatorship	270	270	-	14	84	172
Σ	1254	901	353	274	566	414

Notes: The SVM algorithm produces unbiased measures of democracy only when the priming data reflect the institutional heterogeneity among autocracies and democracies. This table shows how the 1254 autocratically labeled country-years are distributed over different categories. Data comes from Cheibub et al. (2010), Geddes et al. (2014) and Rode and Bjørnskov (2016).

Table 4 Priming data — Institutional heterogeneity among democratic regimes

	Government	**Legislature**		**Voting Procedure**	
	Σ	Unicameral	Bicameral	Majoritarian	Proportional
Parliamentary regime	712	276	436	167	545
Semi-Presidential regime	256	108	148	44	212
Presidential regime	193	62	131	36	157
Σ	1161	446	715	247	914

Notes: The SVM algorithm produces unbiased measures of democracy only when the priming data reflect the institutional heterogeneity among autocracies and democracies. This table shows how the 1161 democratically labeled country-years are distributed over different categories. Data comes from Cheibub et al. (2010), Coppedge et al. (2016) and Rode and Bjørnskov (2016).

presidential regimes, i.e. political systems in which a simple legislative majority is not sufficient to remove the government from office, (ii) parliamentary regimes, i.e. political systems in which a simple legislative majority is necessary to appoint the government and is sufficient to remove the government from office, and (iii) semi-presidential regimes, i.e. political systems in which the government is not elected by the legislature but a simple legislative majority is sufficient to remove the government from office. Column 1 of Table 4 shows that the priming data comprises country-years of each type.

Persson and Tabellini (2005, 2006) suggest that legislative authorities are established through different voting procedures (proportional, majoritarian) and exhibit different structures (unicameral, bicameral). Columns 2 to 5 of Table 4 illustrate that the priming data also reflects these types of institutional heterogeneity.

3.3.3 Alternative selection procedures

The baseline version of the SVM algorithm uses the measures of democraca developed by Pemstein et al. (2010) and Teorell et al. (2016) to select the priming data. We apply a number of alternative selection procedures to check the robustness of the SVM indices. Table 5 reports the results of these analyses. We first use one rather than two measures of democracy for the selection exercise. The resultant changes in the SVM indices are minor and are completely captured by the estimated distribution (see Columns 1 and 2). Next, we increase the number of selection criteria by including the indices published by Freedom House (2015) and Marshall et al. (2016). The SVM indices prove to be robust to these

Table 5 Priming data — Alternative selection procedure

	Continuous SVM Indices					Dichotomous SVM Indices				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Mean Deviation	0.005	0.013	0.008	0.011	0.009	0.005	0.017	0.006	0.016	0.009
Median Deviation	0.004	0.006	0.004	0.004	0.002	0.000	0.000	0.000	0.000	0.000
Correlation	0.999	0.999	0.999	0.999	0.999	0.989	0.966	0.987	0.968	0.982
Coverage Rate	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: To use SVM for the measurement of democracy, we need observations—called *priming data*—with a generally acknowledged degree of democratization. To select country-years that are suited for this purpose, we proceed from two presumptions: (i) the most and least democratic regimes are uncontroversial and thus suitable, and (ii) a country-year belongs to the group of the least (most) democratic regimes when it is in the lower (upper) decile of the index developed by Pemstein et al. (2010), or when it is in the lower (upper) decile of the index developed by Teorell et al. (2016). This table shows how the SVM algorithm responds to a change in the selection procedures. In Column 1, the selection is based on the lower and upper deciles of the indicator developed by Teorell et al. (2016). In Column 2, the selection is based on the lower and upper deciles of the indicator developed by Pemstein et al. (2010). In Column 3, the selection is based on the lower and upper deciles of the indicators developed by Pemstein et al. (2010), Teorell et al. (2016), and Marshall et al. (2016). In Column 4, the selection is based on the lower and upper deciles of the indicators developed by Pemstein et al. (2010), Teorell et al. (2016), and Freedom House (2015). In Column 5, the selection is based on the lower and upper deciles of the indicators developed by Marshall et al. (2016), Pemstein et al. (2010), Teorell et al. (2016), and Freedom House (2015). We report four statistics to summarize the comparison between the baseline indicators and the alternative indicators: (i) the mean of the absolute deviations, (ii) the median of the absolute deviations, (iii) the correlation coefficient, and (iv) the coverage rate, i.e. the share of country-years for which the alternative indicator is between the upper and lower bound of the baseline distribution (CDist, DDist).

extensions (see Columns 3, 4, and 5).

3.4 Reliability of the measures of democracy

We computed measures of democracy for 186 countries and the time period from 1960 to 2014. This section examines twelve of these countries to illustrate the reliability of the SVM indices.[23]

Democratic representation differs across countries. Figure 1 presents the SVM indices of three democracies to illustrate that the SVM algorithm avoids preferential treatment of a particular form of government. The selection includes: the semi-presidential regime in Finland, the parliamentarian regime in New Zealand, and the plebiscitary democracy in Switzerland.[24]

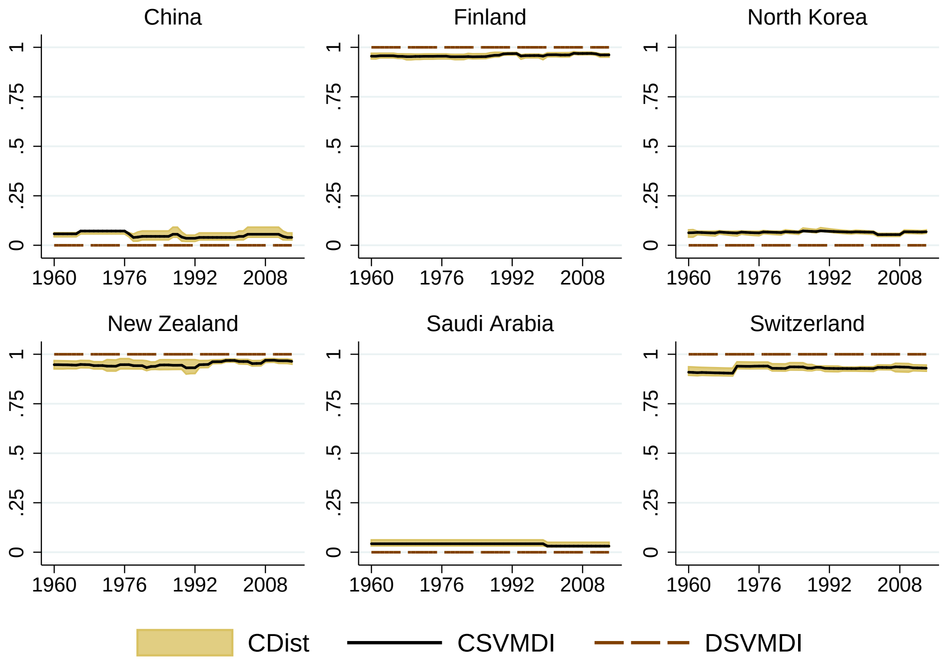
Autocracies are heterogeneous as well. We consider three different types of authoritarianism to demonstrate that the SVM algorithm deals adequately with this institutional heterogeneity. Our selection includes: the royal dictatorship in Saudi Arabia, the military dictatorship in North Korea, and the civil dictatorship in China. Figure 1 shows that the SVM indices do not favor any of these autocratic regimes.

More interesting are, of course, countries that underwent marked changes in regime type, such as Chile. In the period from 1960 to 1973, the multi-party regime in Chile was confronted with growing tension between conservative and left-wing movements. In 1973, the polarization culminated in the removal of the socialist president Salvador Al-

[23] The list of countries includes: Chile, China, Finland, Myanmar (Burma), New Zealand, North Korea, Russia (Soviet Union), Saudi Arabia, South Africa, South Korea, Switzerland, and Turkey. All continuous and dichotomous measures of democracy can be obtained via `http://apps.wiwi.uni-wuerzburg.de/svmdi/`.

[24] The continuous indices are marginally lower for Switzerland than for Finland and New Zealand during the time period from 1960 to 1970. This difference is adequate because of the disfranchisement of Swiss women.

Figure 1 Political Institutions in China, Finland, New Zealand, North Korea, Saudi Arabia, and Switzerland (SVM indices, 1960 – 2014).



Notes: This figure presents the SVM indices of China, Finland, North Korea, New Zealand, Saudi Arabia, and Switzerland. These examples show that the SVM indices are not biased against a specific type of autocracy or democracy. The examples of China and Finland also show that there is not necessarily a difference between those country-years which belong to the priming data and those country-years that are not part of the priming data.

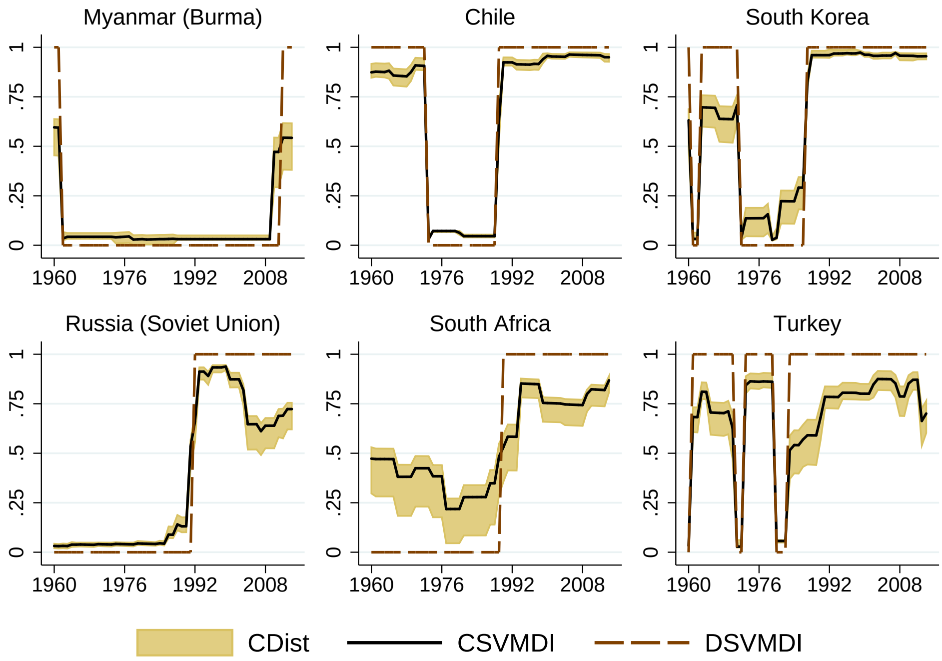
lende through a military coup (Collier and Sater, 2004, Nohlen, 2005). At this point, the SVM indices drop (see Figure 2). After seizing power, the regime led by General Pinochet banned political parties, arrested political opponents, closed both chambers of the parliament, and established an authoritarian constitution. Pinochet's autocratic governance ended in 1989 and all successive governments were legitimized through free elections.

Following its independence from the United Kingdom in 1948, competitive elections were held in Myanmar. This democratic period ended with the military coup of 1962, which ushered in a long period of dictatorship (Nohlen et al., 2001). In 1974, the military government organized parliamentary elections and installed a single-party regime (Devi, 2014). The SVM indices do not react to these events (see Figure 2). We are confident that this persistence is justified because of the absence of electoral competition. In 1990, a multiparty election took place and was won by the National League of Democracy. However, the military government prevented the democratic takeover (Nohlen et al., 2001). In 2008, President Thein Sein signed a new constitutional law which paved the way for the parliamentary election in 2010 (Rieffel, 2013).[25]

The Soviet Union existed from 1922 to 1991 and was a socialist state with a single-

[25] Electoral competition was low in 2010 because: (i) the military appointed one quarter of the parliamentarians, and (ii) the National League of Democracy boycotted this election. The National League of Democracy attended to the by-election in 2012 and received the majority of votes (Rieffel, 2013).

Figure 2 Political Institutions in Chile, Myanmar (Burma), Russia (Soviet Union), South Africa, South Korea, and Turkey (SVM indices, 1960 – 2014).



Notes: This figure presents the SVM indices of Myanmar (Burma), Chile, South Korea, Russia (Soviet Union), South Korea, and Turkey. These examples show that the institutional transitions being indicated by the SVM indices can be justified through historical records.

party government. In 1986, Mikhail Gorbachev, the General Secretary of the Communist Party, introduced two new policies—*Glasnost* and *Perestroika*[26]—thereby initiating an institutional change (Sakwa, 2005). The democratization process continued in the Russian Federation, where the first competitive election was held in 1991 and a new constitution was adopted in 1993. Since 2003, the Russian parliamentary and presidential elections have become increasingly less free (Hale et al., 2004, Nohlen and Stöver, 2010, Sakwa, 2014). In light of these facts, we believe that the SVM indices provide an adequate description of the political development in Russia.

Since it gained its independence in 1910, all South African governments were legitimized through competitive elections. However, until the beginning of the 1990s, black Africans were excluded from the political process (Nohlen et al., 1999). We argue that this racial discrimination justifies the low degree of democratization suggested by the SVM indices.[27] From 1989 to 1993, President Frederick de Klerk launched a number of reforms: he lifted the ban on the African National Congress (ANC), released political prisoners, improved political rights, and initiated negotiations on a new electoral law. This process of reform

[26]Perestroika was an economic program that was introduced to modernize the economy of the Soviet Union. Glasnost comprises a number of political reforms that were implemented to lift restrictions on the freedom of speech and information (Sakwa, 1990).

[27]The reduction in the degree of democratization indicated by the SVM indices during the sixties and seventies can be explained by the growing repression of the extra-parliamentarian opposition and the increasing dominance of the ruling party (Beck, 2013, Nohlen et al., 1999).

led to a free parliamentary election in 1994, which was clearly won by the ANC. All subsequent elections had similar results (Beck, 2013, Nohlen et al., 1999).[28]

In 1960, South Korea introduced a parliamentary regime with a bicameral legislature. Multi-party elections were held in the same year and won by the former main opposition party. The military was unwilling to support the new government and seized power in 1961 (Han, 1974). The SVM indices decline sharply because of this coup. In 1963, South Korea implemented a presidential regime and Park Chung-hee became President. According to Nohlen et al. (2001), the election process was "*fairly democratic*", and thus we have no concerns about the appropriateness of the indicated increase in the degree of democratization. After the declaration of martial law in 1972, President Park established a constitutional law that allowed him to restrict political rights, to rule by decree, and to appoint one-third of the members of the parliament (Chang, 2008). We argue that these reforms undermined democratic practices, and thus justify the decrease in the SVM indices. Since 1987, free and competitive elections have been held regularly in South Korea.[29]

In the aftermath of the military coup in 1960, a parliamentary regime with a bicameral legislature and multi-party elections was implemented in Turkey. However, the attempt to establish democratic institutions met with little success because the government coalitions were unable to end the violence between left- and right-wing movements. This resulted in a military intervention in 1971. Democratic processes were resumed two years later and maintained until 1980 (Nohlen et al., 2001). Given these facts, we are confident that our measures provide an adequate description of the political institutions in Turkey during the 1960s and 1970s. The coup in 1980 differs from the previous interventions for two reasons: (i) all existing parties were dissolved, and (ii) the military government allowed only three newly created parties to participate in the parliamentary election in 1983 (Nohlen et al., 2001). We argue that the latter action explains why the continuous measure gradually increases after the dissolution of the third military government.[30]

3.5 SVM algorithm vs. conventional aggregation methods

The evidence presented in the previous section suggests that the SVM algorithm produces sophisticated measures of democracy. In this section, we compare the performance of different aggregation techniques to illustrate that the SVM algorithm outperforms conventional methods.[31]

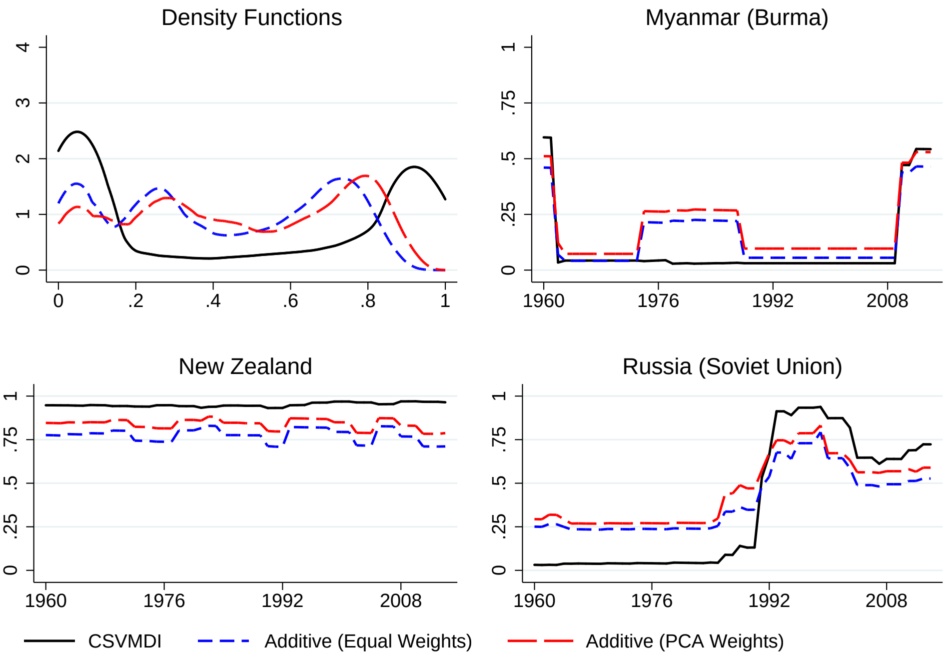
[28]The limited electoral competition that originates from the substantial dominance of the ANC explains why the continuous measure points to some deficiencies.

[29]President Park was assassinated in 1979 and a military coup ousted his successor from power in 1980. The authoritarian regime that followed the coup abdicated its rule in 1987 (Nohlen et al., 2001).

[30]The restriction of party competition applied only to the parliamentary election in 1983. From 1987 to 2014, all parliamentary elections were pluralist and free. The low political integration of the Kurdish citizens is a major reason why our continuous measure indicates some deficiencies.

[31]We need conceptually equivalent measures of democracy to ensure that the results of the comparison reveal the performance differences of the aggregation methods. We apply conventional aggregation methods to our raw data to meet this requirement.

Figure 3 CSVMDI vs. additive measures of democracy



Notes: This figure compares CSVMDI with two additive measures of democracy. All indices are based on the same raw data, i.e. the differences in the degrees of democratization can be attributed to the particular aggregation method. The graph on the upper left compares the density functions of the three indicators. We present three country examples to illustrate that additive aggregation methods create downward-biased indicators for established democracies and upward-biased indicators for specific types of autocracy.

Additive aggregation methods

Additive indices are based on three assumptions: the characteristics affect the degree of democratization independently from one another, the marginal effect of a characteristic is positive and constant, and the degree of democratization is equal to 1 only if all characteristics are equal to 1. In the following, we illustrate that these functional assumptions create systematic malfunctions in the measures of democracy.

Figure 3 compares CSVMDI with two additive measures of democracy.[32] The comparison reveals two notable differences in the density functions. First, the distributions of the additive indicators are trimodal, while the distribution of CSVMDI has only two modes. Second, the upper mode of CSVMDI is located at a higher degree of democratization than the upper modes of the additive indices.

How can we explain these differences? We suspect that the additive indicators are downward-biased for established democracies. Figure 3 presents the observations for New Zealand to substantiate this hypothesis. We observe that the additive indices suggest a

[32]We apply two standard weighting methods to create the additive indices: first, we weight all characteristics equally, and second, we extract the weights from a principal component analysis (PCA). The indicator developed by Dreher (2006) serves as model for the PCA approach. The estimated weights are distributed as follows: $\omega_{Part1} = 0.146$, $\omega_{Part2} = 0.085$, $\omega_{Comp1} = 0.089$, $\omega_{Comp2} = 0.066$, $\omega_{Comp3} = 0.061$, $\omega_{Comp4} = 0.086$, $\omega_{Comp5} = 0.080$, $\omega_{PoRi1} = 0.197$, $\omega_{PoRi2} = 0.190$.

lack of democracy. We argue that this is an odd rating because there is no evidence of disfranchisement, restricted political competition, or limited political discourse. Appendix Figure B.3 illustrates that the same issue applies to other democratic regimes. We can explain these distortions through the assumption that a regime is fully democratic ($d = 1$) only if all characteristics reach their maximum.[33]

Furthermore, in 1974, twelve years after seizing power, the Myanmarese military organized a single-party election. Figure 3 shows that the indicators react differently to this event: while the additive indices increase significantly, CSVMDI remains stable at a low level.[34] We argue that the latter response is more appropriate since a certain degree of political competition is essential for the existence of democratic institutions. Appendix Figure B.4 shows that the same problem applies to other electoral autocracies. This figure also illustrates that an upward distortion is rather unlikely for autocracies in which the government does not hold elections. This pattern can be explained by the assumption that the characteristics influence the degree of democratization independently from one another.[35]

Taken together, all these arguments suggest that additive aggregation methods produce downward-biased indices for democratic regimes and upward-biased indices for certain autocracies. The consequence of these biases is that additive indicators underestimate the strength of institutional changes. In Figure 3, we present the case of Russia to exemplify this underestimation.

Multiplicative aggregation methods

Multiplicative indices assume that the characteristics reflect necessary conditions, the marginal effect of a characteristic is non-negative and is dependent on the other characteristics, and the degree of democratization is equal to 1 only if all characteristics are equal to 1. The consequences of these assumptions are illustrated in the remainder of this section.

Figure 4 suggests that multiplicative aggregation methods do not produce upward-biased indicators for specific autocracies.[36] We also observe that the multiplicative indica-

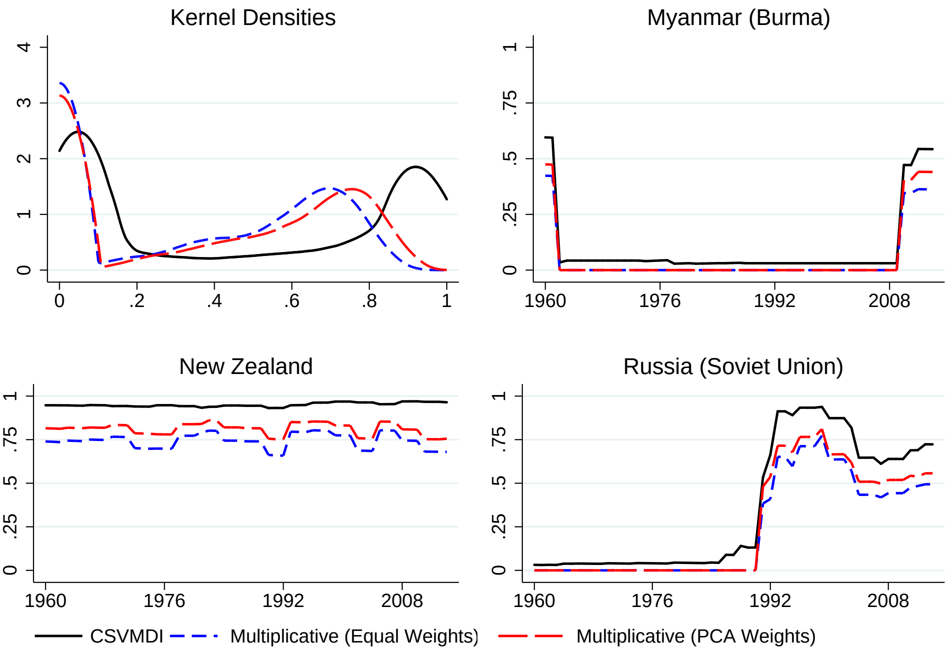
[33]In our case, this assumption implies, for example, that voter turnout needs to be 100 percent and that the number of votes won by the leading party has to be equal to the number of votes won by the leading opposition party. Therefore, if some voters voluntarily decide to abstain from involvement in the election process, or if the leading party is simply more popular than the leading opposition party, the additive indices indicate a lack of democracy. We are not aware of any theory imposing these requirements and thus argue that this assumption creates a downward distortion for democratic regimes.

[34]Figure 3 also shows that, in 1988, the additive indices decrease towards 0. In this year, the military abolished the single-party system (Nohlen et al., 2001).

[35]This assumption implies, in particular, that a higher level of electoral participation increases the degree of democratization, even if there is no political competition. Therefore, it is clear why there is an upward bias for electoral autocracies. We have doubts regarding the logic of this assumption and are not familiar with any theory that could justify it. Furthermore, one may argue that the upward bias only appears because electoral participation is part of the concept of democracy. The example of China illustrates that this argument cannot be supported (see Figure B.4). In this case, the problem arises because the characteristics indicate a limited degree of political discourse.

[36]There is no upward distortion for electoral autocracies due to the assumption that the characteristics reflect necessary conditions. This assumption implies that the degree of democratization is greater than 0 only if all characteristics of democracy are greater than 0. Therefore, an increase in electoral participation does not increase the degree of democratization if there is no political competition.

Figure 4 Comparison between CSVMDI and two multiplicative measures of democracy.



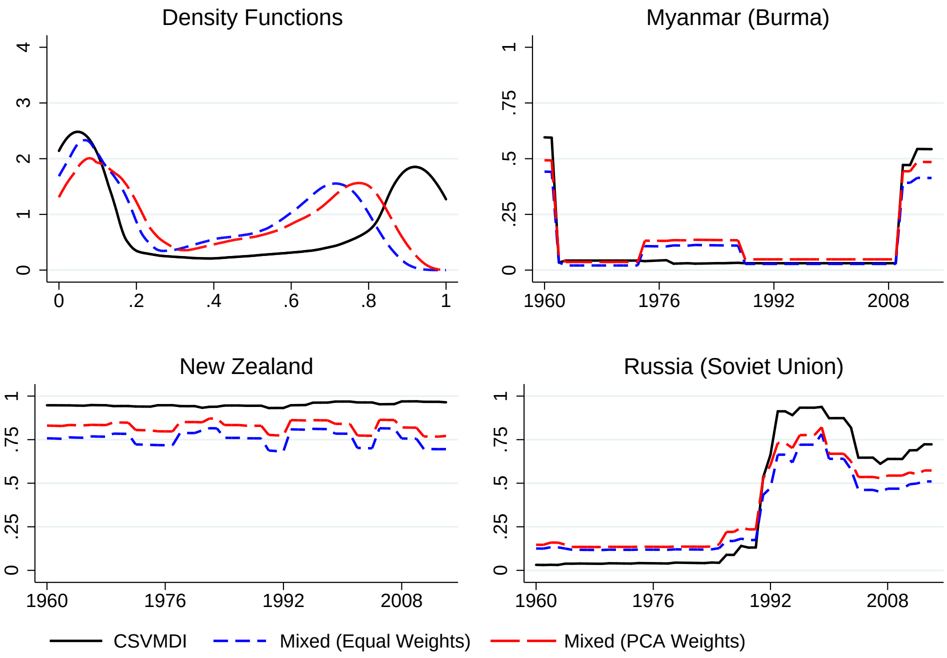
Notes: This figure compares CSVMDI with two multiplicative measures of democracy. All indices are based on the same raw data, i.e. the differences in the degrees of democratization can be attributed to the particular aggregation method. The graph on the upper left compares the density functions of the three indicators. We present three country examples to illustrate that multiplicative aggregation methods create downward-biased indicators for democratic regimes and perform rather well for autocratic regimes.

tors react rather slowly to institutional changes in autocratic regimes. We argue that this limited responsiveness explains the small differences in the density functions at the lower end of the spectrum. To substantiate our argument, we use the example of the Soviet Union in the time period from 1980 to 1990. We observe that the multiplicative indicators remain completely unchanged over time, whereas CSVMDI increases slightly at the end of the decade. The latter reaction seems to be more appropriate in light of the political reforms initiated by Mikhail Gorbachev.

In addition, we observe that the multiplicative indicators are lower than expected for democratic regimes. These downward distortions can still be explained by the assumption that a regime is fully democratic only if all characteristics are equal to 1.

Taken together, the results indicate that multiplicative aggregation methods produce rather reliable indicators for autocratic regimes. However, we observe that the multiplicative measures of democracy are downward-biased for established democracies. We thus conclude that multiplicative indices underestimate the strength of institutional transitions.

Figure 5 Comparison between CSVMDI and two mixed measures of democracy.



Notes: This figure compares CSVMDI with two mixed measures of democracy. All indices are based on the same raw data, i.e. the differences in the degrees of democratization can be attributed to the particular aggregation method. The graph on the upper left compares the density functions of the three indicators. We present three country examples to illustrate that mixed aggregation methods produce downward-biased indicators for democratic regimes and upward-biased indicators for certain autocracies.

Mixed aggregation methods

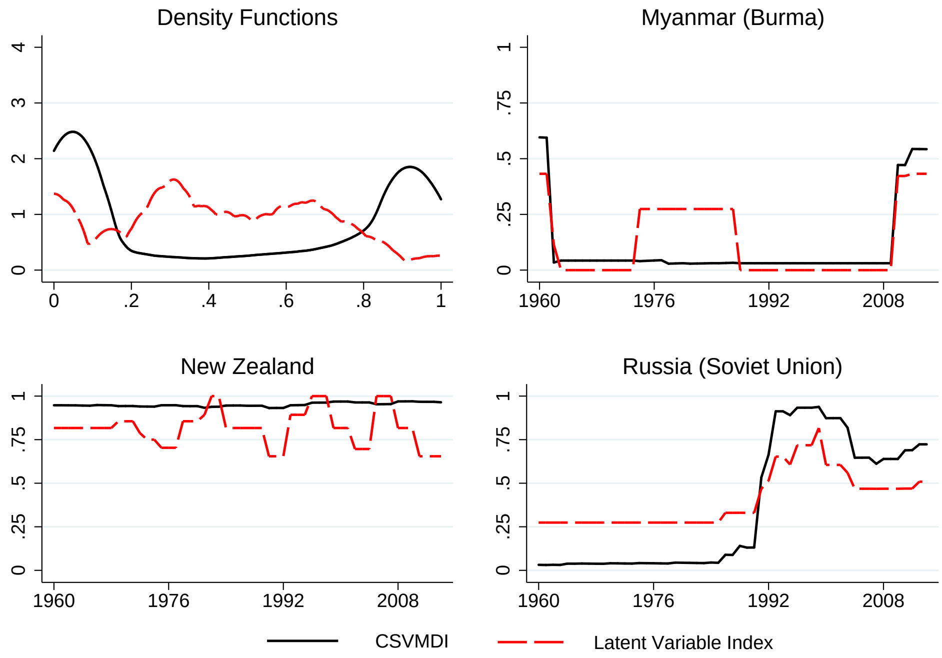
Teorell et al. (2016) take the average of an additive and a multiplicative indicator

$$\mathfrak{F}^{\mathfrak{D}}(\mathbf{x}) \;=\; 0.5 \cdot [\omega_1 x_1 + \ldots + \omega_m x_m + x_1^{\omega_1} \cdot \ldots \cdot x_m^{\omega_m}] \quad \text{with} \quad \omega_i \geq 0, \tag{20}$$

and argue that this "mixed" procedure creates reliable measures of democracy. Their approach is based on three assumptions: the marginal effect of a characteristic on the degree of democratization is positive and is dependent on the other characteristics, the degree of democratization is 0, only when all characteristics are equal to 0, and the degree of democratization is 1 only when all characteristics are equal to 1.

Figure 5 indicates that mixed indices can be downward-biased for highly democratic regimes. This result is not surprising given that there are downward distortions in both additive and multiplicative measures of democracy (see Figures 3 and 4). In addition, we observe small upward distortions for specific autocracies. This latter behavior can be explained by the fact that the multiplicative component does not completely compensate for the measurement error produced by the additive component.

Figure 6 Comparison between CSVMDI and a latent variable index.



Notes: This figure compares CSVMDI with a latent variable index. The indices are based on the same raw data, i.e. the differences in the degrees of democratization can be attributed to the particular aggregation method. The graph on the upper left compares the density functions of the indicators. We present three country examples to illustrate that the latent variable index is downward-biased for democratic regimes and upward-biased for certain autocracies.

Latent variable approach

Pemstein et al. (2010) and Márquez (2016) use a latent variable approach to synthesize additive and multiplicative measures of democracy. The latent variable method proceeds from the assumption that conventional aggregation techniques perform well on average:

$$d^{\mathfrak{D}} = d_j^{\mathfrak{D}} + \varepsilon_j \quad \text{with} \quad \varepsilon_j \sim \mathcal{N}(0, \sigma_j), \tag{21}$$

where $d^{\mathfrak{D}}$ denotes the (unobserved) true degree of democratization, $d_j^{\mathfrak{D}}$ is the degree of democratization that is reported by index j, and σ_j denotes the error variance of index j.

The evidence provided in the previous subsections suggests that the measurement errors produced by conventional aggregation methods are systematic in nature rather than randomly distributed. Therefore, the core assumption of the latent variable approach is unlikely to be satisfied. We suspect that this violation leads to misleading measures of democracy.

We create a latent variable index using the six previously introduced indices. Figure 6 illustrates that the latent variable indicator is upward-biased for specific autocracies and downward-biased for democratic regimes.

Interim conclusion

Conventional aggregation techniques make assumptions about the functional relationship between the characteristics and the degree of democratization. The presented empirical evidence suggests that these assumptions cause systematic measurement distortions at the upper and lower ends of the democracy spectrum. The consequence of these distortions is an underestimation of the strength of institutional transitions. The SVM algorithm does not require functional assumptions and thus produces more adequate measures of democracy.

One concern might be that these findings are caused by a strategic selection of the characteristics. Appendix A.2 presents two supplementary analyses to allay this concern. The first applies an alternative raw data set, i.e. we illustrate that the systematic underestimation does not disappear when using the conceptualization and operationalization developed by Teorell et al. (2016). The second analysis considers four standard measures of democracy, showing that these indicators suffer from the predicted measurement distortions.

Further concerns might be related to the practical importance of measurement issues. We address the question of the empirical consequences in the remainder of this study by investigating the long-run effect of democracy on economic development. We chose this research area because of its topicality and to provide an additional explanation for why empirical studies draw different conclusions about the relationship between democracy and economic development.[37]

4 Democracy and economic development

4.1 Econometric methodology

The question of how to measure the degree of democratization is only one of several challenges encountered when investigating the relationship between democracy and economic development. Another difficulty is that autocratic regimes differ from democratic regimes in cultural, geographical, and historical aspects which also affect the level of GDP per capita. We use a conventional fixed effects model to deal with the heterogeneity in time-invariant factors

$$y_{it} = \sum_{l=1}^{4} \theta_l \, y_{it-l} + \gamma d_{it} + \eta_i + \xi_t + \upsilon_{it}, \tag{22}$$

where y_{it} denotes the log of GDP per capita in country i and year t, d_{it} indicates the degree of democratization, η_i and ξ_t are country and year fixed effects, and υ_{it} is the idiosyncratic error term.

[37] We use the phrase "additional explanation" to indicate that our empirical analysis complements the recent work of Acemoglu et al. (2014), who emphasizes the importance of the econometric methodology.

Papaioannou and Siourounis (2008) and Brückner and Ciccone (2011) illustrate that democratic transitions are often preceded by temporal declines in GDP per capita. Acemoglu et al. (2014) take up on this point as well, arguing that the regression coefficient on democracy (γ) is biased if the income dynamics behind these GDP per capital dips are not adequately modeled. To avoid this problem, we follow Acemoglu et al. (2014) by including four lags of GDP per capita in the regression model.

The impact of democracy on economic development can be identified with the fixed effects approach only if the idiosyncratic error term is uncorrelated with the past, current, and future realizations of the explanatory variables. This condition may be violated because of omitted dynamic factors (Durlauf et al., 2005). We address this endogeneity issue with a standard two-stage least squares (2SLS) approach, in which the first lag of the average degree of democratization in neighboring countries serves as the instrument for the domestic level of democracy (Acemoglu et al., 2014, Aidt and Jensen, 2014, Giuliano et al., 2013, Persson and Tabellini, 2009):

$$d_{it} \;=\; \sum_{l=1}^{4} \rho_l \, y_{it-l} + \beta Z_{it-1} + \eta_i + \xi_t + \varepsilon_{it}.^{38} \tag{23}$$

The motivation for this instrumentation strategy is that institutional transitions often occur in regional waves, and thus a strong correlation exists between the regional and the domestic degree of democratization (Gleditsch and Ward, 2006, Huntington, 1993).

4.2 Baseline estimation

The results of the baseline estimation are presented in Column 1 of Table 6. The underlying sample comprises 170 countries and covers the time period from 1960 to 2014. The standard errors are clustered at the country level and robust to heteroscedasticity. The data on GDP per capita comes from the Penn World Tables and the degree of democratization is measured by the continuous version of the SVM index.

The baseline estimate suggests that democracy has a positive effect on economic development. The regression coefficient is statistically significant at the 10 percent level. Put differently, the transition from autocracy ($d = 0$) towards democracy ($d = 1$) leads to an increase in GDP per capita by about two percent.

The first-stage result is reported in Panel B. The estimate reveals that the regional degree of democratization is positively correlated with the domestic degree of democratization. The first-stage coefficient is 0.6344 and statistically significant at the 1 percent

[38] Let $\mathcal{R} = \{1, ..., R\}$ denote a set of regions, where each country i belongs exactly to one region r, and assume that N_{rt} is the number of countries in region r at period t. In this case, the degree of democratization in neighboring countries—i.e. our instrumental variable Z^r_{it}—is calculated by

$$Z^r_{it} = \frac{1}{N_{rt} - 1} \sum_{\{j \neq i | r' = r, r' \in \mathcal{R}\}} d_{jt},$$

where the exclusion of country i in the sum satisfies the exclusion restriction. Table C.2 shows how we divide the world into different regions and how countries are assigned to these regions.

Table 6 Democracy and economic development

	Full sample		Restricted sample		
	CSVMDI	DSVMDI	CSVMDI	CSVMDI	DSVMDI
	(1)	(2)	(3)	(4)	(5)
	Panel A: Second-stage estimates				
Democracy	0.0198*	0.0212*	0.0362***	0.0392***	0.0401**
	(0.0107)	(0.0123)	(0.0125)	(0.0138)	(0.0168)
Log GDP^{pc} $(t-1)$	1.169***	1.169***	1.182***	1.170***	1.170***
	(0.0315)	(0.0316)	(0.0343)	(0.0333)	(0.0335)
Log GDP^{pc} $(t-2)$	-0.189***	-0.187***	-0.190***	-0.190***	-0.188***
	(0.0269)	(0.0272)	(0.0317)	(0.0313)	(0.0319)
Log GDP^{pc} $(t-3)$	0.0428	0.0432	0.0351	0.0368	0.0372
	(0.0303)	(0.0304)	(0.0307)	(0.0301)	(0.0306)
Log GDP^{pc} $(t-4)$	-0.0577***	-0.0585***	-0.0595***	-0.0540***	-0.0551***
	(0.0141)	(0.0142)	(0.0138)	(0.0131)	(0.0132)
Investment $(t-1)$				0.0354*	0.0324
				(0.0198)	(0.0204)
Inflation $(t-1)$				-0.0335**	-0.0343**
				(0.0159)	(0.0175)
Openness $(t-1)$				0.0181**	0.0181**
				(0.00709)	(0.00717)
	Panel B: First-stage estimates				
Democracy (Region)	0.6344***	0.5104***	0.6034***	0.5880***	0.4462***
	(0.0726)	(0.0820)	(0.0806)	(0.0839)	(0.0939)
Observations	7,606	7,606	6,832	6,832	6,832
Countries	170	170	169	169	169
CD F-Stat.	844.79	403.12	612.69	558.74	263.70
SY 10% IV size	16.38	16.38	16.38	16.38	16.38
AR p-value	0.06	0.06	0.00	0.00	0.00
SW p-value	0.04	0.04	0.00	0.00	0.00
OP F Stat.	111.72	96.86	93.38	97.90	84.36
OP τ =5%	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.002	0.000	0.016	0.018	0.013
UCI (upper)	0.038	0.044	0.056	0.060	0.067
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000
Equal. (p-value)	–	0.905	–	0.823	0.815
R-Squared	0.96	0.96	0.96	0.96	0.96

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. Panel A reports second-stage estimates. Panel B presents first-stage estimates. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects and year fixed effects. The bottom section of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP $\tau = 5\%$. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). Equal (p-value) refers to the p-value of the Wald test for equality of parameter estimates. The bases of comparison are the point estimates reported in Column (1) and Column (3). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**p < .05$, $***p < .01$.

level. Furthermore, the F-statistic of the first-stage regression suggests that the instrumental variable is not weak. This conclusion is supported by the results of the weak-instrument tests developed by Anderson and Rubin (1949), Kleibergen and Paap (2006), Olea and Pflüger (2013), and Stock and Wright (2000).

The baseline estimation assumes that—conditional on the control variables—the first lag of the regional degree of democratization influences the level of GDP per capita only through its effect on the domestic degree of democratization. An argument for why this exclusion restriction might be violated is that changes in the regional degree of democratization might have an effect on the regional level of political instability, and thus affect domestic prices, investments, and trade flows. To address this concern, we augment the regression model, controlling for the investment share, the inflation rate, and the degree of trade openness. The data is taken from the Penn World Tables and the World Bank. We expect that the model extension gives rise to a change in the estimate of the effect of democracy on economic development only when the importance of the alternative mechanism is not negligibly small.

Extending the regression model decreases the sample size by about 800 observations. We therefore run the baseline regression first on the country-years for which we have information on investment, inflation, and trade openness. The estimation results are presented in Column 3 of Table 6 and illustrate that the second-stage estimate of the effect of democracy increases when we restrict the sample.[39]

The consequences of the model extension are illustrated in Column 4 of Table 6. We observe that the regression coefficient remains almost unchanged when we control for the share of investments, the inflation rate, and the degree of trade openness (p-value of the Wald test: 0.823). The robustness of the estimate suggests that the regional degree of democratization does not exert influence on per capita GDP through domestic prices, investments, and trade flows.

There may be other unobserved channels that could violate the exclusion restriction. We extend the regression model in two different ways to allay this concern. In the first extension, we control for human capital, public health, civil conflict, natural disasters, and government consumption.[40] Appendix Table C.3 illustrates that the inclusion of these control variables does not significantly change the regression result. Second, we control for socioeconomic factors that are correlated within regions and might affect both the degree of democratization and per capita GDP . Table C.4 in the Appendix indicates that the estimated coefficient remains stable when we control for the regional level of investment, inflation, trade openness, human capital, public health, civil conflict, natural disaster, and government consumption. We also perform the Union of Confidence Interval (UCI) test developed by Conley et al. (2012). The UCI test proceeds from the assumption that the exclusion restriction is not completely satisfied and calculates an interval estimate which takes this violation into account. The results of the UCI test are reported at the bottom of Table 6. We observe that the lower bound of the interval estimate is non-negative in all model specifications, suggesting that the positive effect of democracy on economic

[39] The increase in the estimated coefficient suggests that there is some heterogeneity in the effect of democracy on economic development. Investigating the causes of the heterogeneity is beyond the scope of this study, however it might be an interesting topic for future research.

[40] The data on human capital and government consumption is taken from the Penn World Tables (Feenstra et al., 2015). We use data on life expectancy from the World Bank (2017) to control for the level of public health. Information about natural disasters is provided by the International Disaster Database (Guha-Sapir et al., 2015). The index developed by Bluhm et al. (2016) is used to measure civil conflict.

development is not the consequence of an invalid instrumental variable.

4.3 Robustness checks

Table 7 presents the results of several robustness checks. In Columns 1 and 2, the dependent variable is the growth rate of GDP per capita rather than the log of GDP per capita. The regression coefficients support the hypothesis that there is a positive and statistically significant effect of democracy on economic development.

A number of economists argue that annual data is not suitable for analyzing the determinants of economic development. These scholars prefer data that is averaged over multiple years because this approach may filter out business cycle fluctuations and may mitigate the influence of measurement uncertainty (Durlauf et al., 2005). Columns 3 and 4 of Table 7 suggest that the positive effect of democracy on economic development persists when we use five-year data averages rather than annual data.

The SVM algorithm creates for each country-year a distribution of indicators that can be used to gauge the degree of measurement uncertainty (for details, see Sections 3.1 and 3.2). This feature has so far not been exploited in the empirical analysis; the robustness of the baseline results could thus be challenged. To address this concern, we apply the multiple imputation technique developed by Rubin (1996, 2004). This procedure consists of four stages: (i) random extraction of degrees of democratization from the distributions, (ii) estimation of the empirical model, (iii) multiple repetition of the first two stages multiple times, and (iv) averaging over all regression coefficients to calculate the effect of democracy on economic development. Columns 5 and 6 of Table 7 illustrate that the baseline results hold when acknowledging the uncertainty in the measure of democracy.

In the baseline regression, the instrumental variable is based on a classification scheme that assigns each country to one out of sixteen world regions (see Table C.2). We also follow Acemoglu et al. (2014) in using the classification scheme of the World Bank to divide the world into seven rather than sixteen geographical regions.[41] The estimation results are reported in Columns 7 and 8 of Table 7. In both cases, we observe that the effect of democracy on economic development continues to be positive. However, the regression coefficient is statistically insignificant when we consider the full sample. We argue that this result is not surprising, given that the statistical tests developed by Anderson and Rubin (1949) and Stock and Wright (2000) suggest a weak instrument.[42]

To be consistent with the modeling of the dynamics in income per capita, Acemoglu et al. (2014) use the first four lags of the regional degree of democratization as instrumental variables. Columns 9 and 10 of Table 7 display that the second-stage estimate is positive and statistically significant when we apply this alternative instrumentation strategy.

[41] These seven geographical regions are: East Asia and the Pacific, Europe and Central Asia, Latin America and the Caribbean, the Middle East and North Africa, North America, South Asia, and Sub-Sahara Africa.

[42] The weak instrument problem arises because the classification of regions is determined too broadly. The classification scheme of the World Bank consolidates, for example, countries from Europe and Central Asia into the same group. The substantial geographical distance contradicts the argument of democratization waves developed by Huntington (1993) and Gleditsch and Ward (2006).

Table 7 Democracy and economic development — Robustness econometric methodology (CSVMDI)

	Growth Rate of GDP^{pc}		Five-Year Data		Multiple Imputation		World Bank Regions		Instrument (4 Lags)		Difference GMM		Placebo	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
CSVMDI	0.0288**	0.0466***	0.135*	0.246***	0.0196*	0.0389***	0.0234	0.0480**	0.0212**	0.0408***	0.0198*	0.0310*	0.0584	0.0808
	(0.0128)	(0.0159)	(0.0690)	(0.0921)	(0.0106)	(0.0136)	(0.0185)	(0.0186)	(0.0104)	(0.0137)	(0.0101)	(0.0162)	(0.0651)	(0.0707)
Log GDP^{pc} $(t-1)$	-0.0233	-0.0167	0.849***	0.825***	1.169***	1.171***	1.169***	1.170***	1.169***	1.170***	1.418***	0.942***	1.180***	1.168***
	(0.0407)	(0.0339)	(0.0247)	(0.0261)	(0.0315)	(0.0333)	(0.0314)	(0.0334)	(0.0315)	(0.0333)	(0.102)	(0.178)	(0.0342)	(0.0332)
Log GDP^{pc} $(t-2)$	0.0437	0.0343			-0.189***	-0.190***	-0.188***	-0.189***	-0.189***	-0.190***	-0.446***	0.102	-0.188***	-0.185***
	(0.0403)	(0.0392)			(0.0269)	(0.0313)	(0.0272)	(0.0315)	(0.0315)	(0.0313)	(0.129)	(0.190)	(0.0318)	(0.0315)
Log GDP^{pc} $(t-3)$	-0.0360	-0.0363			0.0429	0.0368	0.0429	0.0367	0.0429	0.0368	0.0778**	-0.00703	0.0349	0.0364
	(0.0224)	(0.0239)			(0.0303)	(0.0301)	(0.0302)	(0.0302)	(0.0303)	(0.0301)	(0.0363)	(0.0365)	(0.0312)	(0.0308)
Log GDP^{pc} $(t-4)$	-0.0270	-0.0260			-0.058***	-0.054***	-0.058***	-0.054***	-0.058***	-0.054***	-0.052**	-0.040**	-0.059***	-0.053***
	(0.0196)	(0.0195)			(0.0141)	(0.0131)	(0.0142)	(0.0130)	(0.0141)	(0.0131)	(0.0228)	(0.0185)	(0.0152)	(0.0142)
Investment $(t-1)$		0.0294		0.348***		0.0354*		0.0337*		0.0351*		0.163***		0.0274
		(0.0234)		(0.136)		(0.0198)		(0.0196)		(0.0199)		(0.0601)		(0.0261)
Inflation $(t-1)$		-0.0181*		-0.980*		-0.0035**		-0.0341**		-0.0336**		0.0268		-0.0367*
		(0.0106)		(0.325)		(0.0159)		(0.0166)		(0.0160)		(0.0379)		(0.0204)
Openness $(t-1)$		0.0156**		0.0791		0.0181**		0.0188**		0.0182**		-0.0023**		0.0216***
		(0.00688)		(0.0512)		(0.00708)		(0.00743)		(0.00711)		(0.00109)		(0.00814)
Observations	7,436	6,664	1,489	1,371	7,606	6,832	7,606	6,832	7,606	6,822	7,436	6,662	7,606	6,832
Countries	170	170	170	169	170	169	170	169	170	169	170	169	170	169
CD F-Stat.	806.99	535.57	136.50	92.76	–	–	305.75	351.87	213.89	143.12	–	–	30.32	27.85
SY 10% IV size	16.38	16.38	16.38	16.38	–	–	16.38	16.38	16.38	16.38	–	–	16.38	16.38
AR p-value	0.02	0.00	0.05	0.00	–	–	0.20	0.00	0.05	0.01	–	–	0.30	0.12
SW p-value	0.02	0.00	0.05	0.00	–	–	0.13	0.00	0.09	0.01	–	–	0.26	0.10
OP F-Stat.	110.19	95.96	107.52	88.72	–	–	60.67	97.90	80.42	100.77	–	–	0.49	0.15
OP τ =5%	33.11	33.11	33.11	33.11	–	–	33.11	33.11	33.11	33.11	–	–	33.11	33.11
UCI (lower)	0.002	0.002	-0.098	0.112	–	–	-0.007	0.020	0.002	0.011	–	–	-0.019	-0.004
UCI (upper)	0.048	0.091	0.376	0.452	–	–	0.054	0.076	0.038	0.078	–	–	0.059	0.083
KP rk LM p-value	0.000	0.000	0.000	0.000	–	–	0.000	0.000	0.000	0.000	–	–	0.098	0.111
R-Squared	0.04	0.04	0.76	0.75	–	–	0.96	0.96	0.96	0.96	–	–	0.96	0.96
Hansen p-value	–	–	–	–	–	–	–	–	–	–	0.072	0.179	–	–
AR(1) p-value	–	–	–	–	–	–	–	–	–	–	0.000	0.064	–	–
AR(2) p-value	–	–	–	–	–	–	–	–	–	–	0.201	0.112	–	–
Instruments	–	–	–	–	–	–	–	–	–	–	120	147	–	–

Notes: Using data from Feenstra et al. (2015), the dependent variables are the growth rate of GDP per capita (Columns 1 and 2) and the log of GDP per capita (Columns 3 – 14). The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. Table reports second-stage estimates (Columns 1 – 10, 13, and 14) and Difference GMM estimates (Columns 11 and 12). Estimates reported in Column 3 and 4 are based on five-year data averages rather than annual data observations. In Columns 1 – 8, the instrumental variable is the regional degree of democratization in period $t-1$ (jackknifed). In Columns 9 and 10, the instrumental variables are the first four lags of the regional degree of democratization (jackknifed). In Columns 1 – 6, 9 and 10, the classification of regions presented in Table C.2 is used to construct the instrumental variables. The classification proposed by the World Bank is used in Columns 7 and 8. In Columns 13 and 14, countries are randomly assigned to regions. The multiple imputation method developed by Rubin (1996, 2004) is used to compute the estimates in Columns 5 and 6. The number of imputation is set to 100. All regressions include country fixed effects and period fixed effects. The bottom part of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP $\tau = 5\%$. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). Hansen p-value refers to the p-value of the overidentification test developed by Hansen (1982). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**\,p < .05$, $***p < .01$.

Furthermore, we implement the Difference GMM estimator developed by Arellano and Bond (1991) to examine whether the usage of internal rather than external instruments leads to significant changes in the estimation results. Columns 11 and 12 of Table 7 show that there are no notable changes in the coefficient estimates.

Finally, we perform placebo tests, in which we randomly assign countries to regions. The estimation results are presented in Columns 13 and 14 of Table 7. As expected, we find that the instrumental variable is unacceptably weak. We also observe that the second-stage estimate is statistically insignificant, suggesting that the baseline result is unlikely to be driven by spurious factors.

4.4 Does the measure of democracy matter?

We now turn to the last and most important part of the empirical analysis, in which we examine how the baseline results react to changes in the measurement of democracy. We investigate the consequences of: (i) using a dichotomous rather than a continuous indicator, (ii) utilizing an alternative concept of democracy, and (iii) applying conventional aggregation techniques instead of the SVM algorithm.

Continuous vs. dichotomous measures of democracy

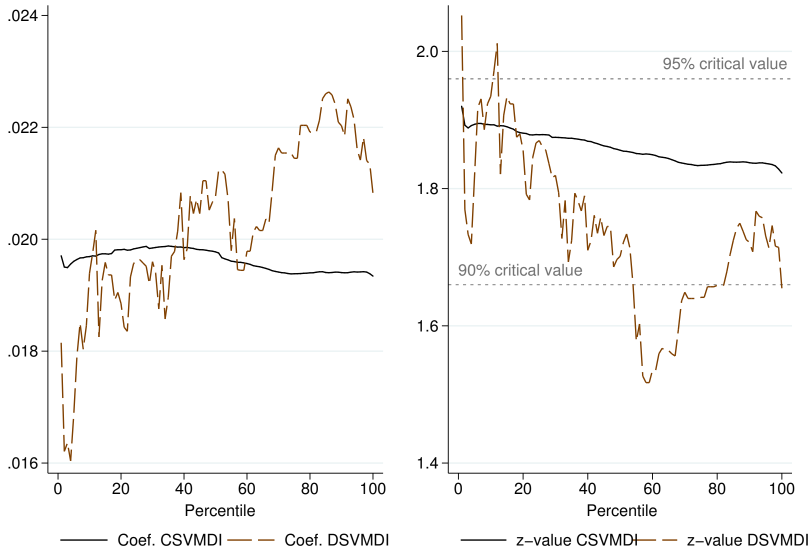
Social scientists still disagree on whether continuous or dichotomous indices are more appropriate for the measurement of democracy. We refrain from taking sides since both approaches offer their own advantages: while continuous measures allow for a finer distinction between democratic regimes, dichotomous indices provide a clearer indication of institutional transitions (Acemoglu and Robinson, 2006, Collier and Adcock, 1999).

In the literature on the determinants of economic development, only two studies address the consequences of replacing a continuous measure with a dichotomous index. Their findings are ambivalent: Doucouliagos and Ulubaşoğlu (2008) suggest that the coefficient of the democracy variable decreases when using a dichotomous index, whereas Cheibub et al. (2010) report the opposite result. These studies compare, however, measures of democracy that differ with regard to their conceptualizations and operationalizations. We are able to avoid this problem because the SVM algorithm produces continuous and dichotomous indicators that are conceptually equivalent (see Section 3.1).

We again carry out the baseline regressions, this time with DSVMDI to provide an idea of how the estimates respond when we use a dichotomous rather than a continuous index. The findings are presented in Columns 2 and 5 of Table 6 and suggest that there are only minor changes in the estimation results for the effect of democracy on GDP per capita.

Do these results allow for the conclusion that the choice of the scale is of little importance? Alvarez et al. (1996) argue that a serious comparison between a continuous and a dichotomous measure of democracy needs to take into account the differences in the type of measurement error: while the continuous index contains a large number of small mistakes, the dichotomous index contains a small number of large mistakes.

Figure 7 Consequences of measurement uncertainty — CSVMDI vs. DSVMDI



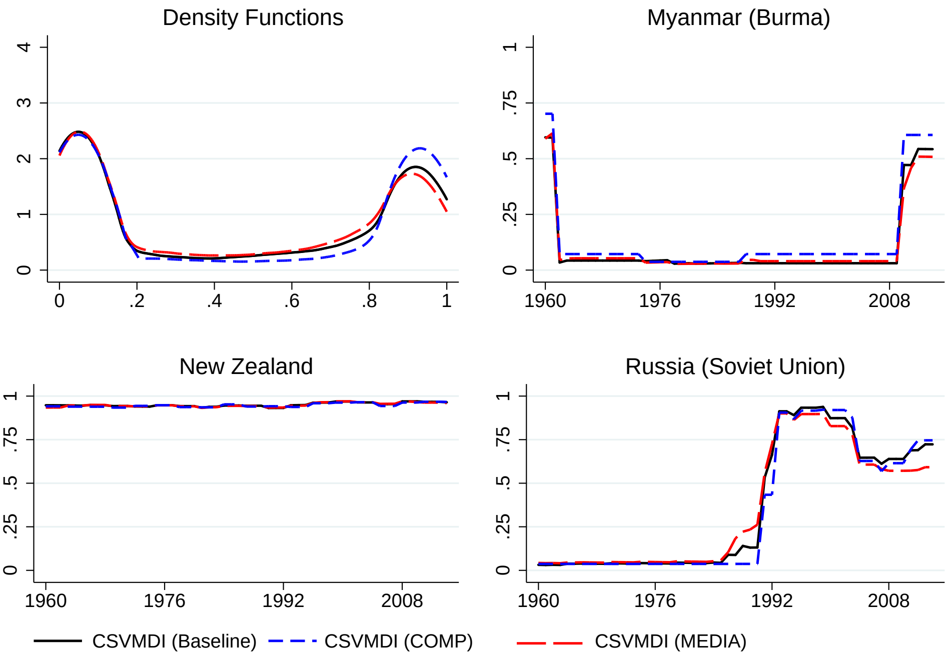
Notes: This figure investigates how the estimation results change when we take into account the measurement uncertainty in continuous and dichotomous measures of democracy. For this purpose, we use the distributions of the SVM indices (CDist, DDist) and estimate the baseline model for every percentile of the distributions. The left graph reports the point estimates of the second-stage regressions, while the right graph presents the levels of statistical significance. All regressions control for the first four lags of the log of GDP per capita, country fixed effects, and year fixed effects. The dependent variable is the log of GDP per capita. The regional degree of democratization in $t-1$ serves as the instrumental variable. The number of observations is equal to 7606.

The remainder of this section investigates whether these different types of uncertainty have different consequences. For this purpose, we use the distributions of the two indicators (CDist, DDist) and estimate the regression model for every percentile of the distributions. The results of this analysis are summarized in Figure 7. We observe that the coefficient estimate is highly volatile when we use the dichotomous measure of democracy: in some instances the estimate is about 1.6 percent and statistically significant at the 5 percent level, whereas in other cases, the estimate is about 2.1 percent and statistically insignificant. By contrast, the regression result is relatively stable when using the distribution of the continuous indicator.

The supplementary material presents additional analyses to illustrate that the pattern observed in Figure 7 can be considered representative. We show in particular that the result holds when using the restricted sample and include control variables in the regression model (see Appendix Figure B.7), and also when using the Difference GMM estimator rather than the 2SLS estimator (see Appendix Figure B.8).

Taken these results together, we are able to conclude that the scaling of the measure of democracy is not irrelevant. We therefore advise scholars to rethink their choice. In contentious cases, the best strategy is to apply both a continuous and a dichotomous indicator. Furthermore, we argue that it is important to acknowledge the uncertainty that exists in every measure of democracy, since not doing so might lead to incorrect

Figure 8 Alternative concepts of democracy



Notes: This figure compares measures of democracy that differ with regard to the concept of democracy. The baseline concept requires the existence of political competition, electoral participation, and political discourse. The concept "COMP" only requires the existence of political competition. The concept "MEDIA" requires the existence of political competition, electoral participation, political discourse, and media independence. The graph on the upper left compares the density functions of the conceptually different indicators. We present three country examples to illustrate that differences in the measures of democracy mainly exist for hybrid regimes.

conclusions about the consequences of democratization. A key advantage of the SVM algorithm is thus that unlike conventional aggregation methods, it provides opportunities to examine the roles of scaling issues and measurement uncertainty.

An alternative concept of democracy

The baseline concept of democracy requires the existence of electoral participation, political competition, and political discourse. In this section, we change the definition. The motivation is twofold: First, a lack of consensus about how to define the term "democratic regime" which requires us to show how the SVM algorithm reacts to conceptual adjustments. Secondly, we know little about the empirical consequences that result from changes in the concept of democracy.

We consider two alternative concepts of democracy. The first is the definition that goes back to Schumpeter (1942), who defines democracy as an "*institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote.*" Schumpeter's definition thus rests on competition for political leadership. The first alternative of the SVM index is therefore computed by using only the five characteristics of political competition. We use the

Table 8 Democracy and economic development — Alternative concepts of democracy

	Full sample			**Restricted sample**		
	Baseline	COMP	MEDIA	Baseline	COMP	MEDIA
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Second-stage estimates					
CSVMDI	0.0198*	0.0208*	0.0186*	0.0392***	0.0406**	0.0390***
	(0.0107)	(0.0117)	(0.0110)	(0.0138)	(0.0155)	(0.0141)
Observations	7,606	7,606	7,606	6,832	6,832	6,832
Countries	170	170	170	169	169	169
CD F-Stat	844.79	625.06	850.51	558.74	385.55	562.00
SY 10% IV size	16.38	16.38	16.38	16.38	16.38	16.38
AR p-val	0.06	0.06	0.08	0.00	0.00	0.00
SW p-val	0.04	0.04	0.07	0.00	0.00	0.00
OP F Stat	111.72	95.31	113.85	97.90	80.41	100.75
OP τ =5%	33.11	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.002	0.000	0.002	0.018	0.021	0.019
UCI (upper)	0.038	0.041	0.035	0.060	0.067	0.059
KP rk LM p-val	0.000	0.000	0.000	0.000	0.000	0.000
Wald Stat. p-value	–	0.929	0.915	–	0.815	0.986
R-Squared	0.96	0.96	0.96	0.96	0.96	0.96
	Panel B: Second-stage estimates					
DSVMDI	0.0212*	0.0188*	0.0194*	0.0401**	0.0376**	0.0353**
	(0.0123)	(0.0114)	(0.0109)	(0.0168)	(0.0159)	(0.0138)
Observations	7606	7606	7606	6832	6832	6832
Countries	170	170	170	169	169	169
CD F-Stat.	403.12	381.80	475.92	263.70	215.29	304.58
SY 10% IV size	16.38	16.38	16.38	16.38	16.38	16.38
AR p-value	0.06	0.08	0.06	0.00	0.00	0.00
SW p-value	0.04	0.06	0.04	0.00	0.00	0.00
OP F Stat.	96.86	80.46	105.68	84.36	87.94	94.49
OP τ =5%	33.11	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.000	-0.001	-0.001	0.013	0.018	0.017
UCI (upper)	0.044	0.036	0.035	0.067	0.058	0.058
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000	0.000
Equal. (p-value)	–	0.833	0.871	–	0.877	0.727
R-Squared	0.96	0.96	0.96	0.96	0.96	0.96

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. The baseline concept of democracy used in Columns 1 and 4 requires the existence of political competition, electoral participation, and political discourse. The concept "COMP" used in Columns 2 and 5 only requires the existence of political competition. The concept "MEDIA" used in Columns 3 and 6 requires the existence of political competition, electoral participation, political discourse, and media independence. Panel A reports second-stage estimates for the continuous indicators and Panel B presents second-stage estimates for the dichotomous indicators. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects, year fixed effects and four lags of the dependent variable. Columns 4 – 6 also include control variables for investment, trade openness, and inflation. The bottom part of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP τ = 5%. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). Equal (p-value) refers to the p-value of the Wald test for equality of parameter estimates. The bases of comparison are the point estimates reported in Column 1 and Column 4. The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**p < .05$, $***p < .01$.

subscript "COMP" to refer to this measure of democracy.

The second concept follows some theorists' argument that a democratic regime requires independence of the media (Schultz, 1998). We add this dimension to the baseline concept and operationalize it with the help of expert-based ratings of media censorship and the harassment of journalists. Both of these ratings are reported in Coppedge et al. (2016). We use the subscript "MEDIA" to refer to this conceptually broader version of the SVM index.

Figure 8 compares these alternative versions of the continuous SVM index. We observe substantial similarities between the three measures of democracy, especially with regard to autocratic regimes (e.g. Myanmar, Soviet Union) and established democracies (e.g. New Zealand). Differences between the indicators exist mainly for hybrid regimes. For example, the level of democracy of the Russian Federation is less pronounced when we consider CSVMDI_{Media}. This difference is plausible because the freedom of the press is more restricted in Russia than are other aspects of democracy (Becker, 2004, Oates, 2007).

We carry out the baseline estimations using the alternative measures of democracy to illustrate how the regression estimates react to a change in the concept of democracy. The results of this analysis are reported in Table 8. Two findings are particularly relevant: (i) The effect of democracy on economic development is positive and statistically significant regardless of the assumed concept of democracy. (ii) The differences in the estimated coefficients are minor, suggesting that conceptual aspects are of minor empirical relevance.[43]

Alternative aggregation methods

In Section 3.5, we arrived at the conclusion that conventional aggregation techniques can produce misleading measures of democracy. In particular, we provided evidence for the hypothesis that conventionally produced indicators underestimate transitions. In this section, we illustrate the resulting empirical consequences.

Consider a simplified version of our regression model in which the level of economic development is a linear function of the degree of democratization:

$$y_t \;=\; \eta + \gamma \cdot d_t + \varepsilon_t \quad \text{with} \quad \eta \geq 0 \text{ and } \gamma \geq 0.^{44} \tag{24}$$

Furthermore, suppose that the level of democracy is low in earlier periods and high in later periods:

$$d_1 \;=\; \ldots \;=\; d_\tau \;<\; d_{\tau+1} \;=\; \ldots \;=\; d_T \quad \text{with} \quad \tau \in \{1, \ldots, T-1\}, \tag{25}$$

where $\tau + 1$ denotes the transition year. Assume further that there is an instrumental variable which allows endogeneity issues—such as reversed causality, omitted factors, and

[43]The supplementary material shows that these results hold when taking into account the measurement uncertainty (see Appendix Figure B.9), and also when using the Difference GMM estimator rather than the 2SLS estimator (see Appendix Figure B.10).

[44]For two reasons, we do not include lags of the dependent variable in the regression model: First, the econometric explanation that we provide also applies when covariates are included. Second, we believe that the argumentation is much clearer when a static model is used instead of a dynamic model. For the same reasons, we consider a simple time-series model rather than a panel model with country fixed effects.

random noise in the explanatory variable—to be dealt with:

$$z_1 \;=\; \ldots \;=\; z_\tau \;<\; z_{\tau+1} \;=\; \ldots \;=\; z_T \quad \text{with} \quad \tau \in \{1, \ldots, T-1\}. \tag{26}$$

Under these assumptions the 2SLS estimator is consistent (Angrist and Pischke, 2009), and some algebra reveals that the 2SLS estimator can be written as

$$\hat{\gamma}_{2SLS} \;=\; \frac{\sum_{t=1}^{T}(y_t - \bar{y})(z_t - \bar{z})}{(d_T - d_1)(z_T - z_1)\left[\tau\left(\frac{T-\tau}{T}\right)^2 + (T-\tau)\left(\frac{\tau}{T}\right)^2\right]} \;\xrightarrow{P}\; \gamma. \tag{27}$$

Suppose now that the measure of democracy is downward-biased for democratic regimes, i.e. the observed degrees of democratization $\{\delta_t\}_{t=1,\ldots,T}$ are correct in the early periods but are too small in the later periods:

$$d_s \;=\; \delta_s \;<\; \delta_t \;<\; d_t \quad \text{with} \quad s \in \{1, \ldots, \tau\} \text{ and } t \in \{\tau+1, \ldots, T\}. \tag{28}$$

Equation (27) implies that the 2SLS estimate increases and may become inconsistent because of this systematic distortion:

$$\delta_T - \delta_1 < d_T - d_1 \;\Rightarrow\; \frac{\sum_{t=1}^{T}(y_t - \bar{y})(z_t - \bar{z})}{(\delta_T - \delta_1)(z_T - z_1)\left[\tau\left(\frac{T-\tau}{T}\right)^2 + (T-\tau)\left(\frac{\tau}{T}\right)^2\right]} \;\xrightarrow{P}\; \tilde{\gamma} \;\geq\; \gamma.\text{[45]} \tag{29}$$

Table 9 reports estimates for the relationship between democracy and economic development and shows how these estimates react to a change in the aggregation technique. Column 1 presents the baseline estimates, while Column 2 reports the estimation results when using the additive index (PCA weights) that is conceptually equivalent with CSVMDI (for details, see Section 3.5). We observe an increases by 36 percent in the size of the estimate and a decrease in the level of statistical significance.[46] In Column 3, the estimated effect is based on the multiplicative indicator (PCA weights). The regression coefficient turns out to be of similar size as in the additive specification. However, in contrast to Column 2, the effect is statistically significant at the 5 percent level. Column 4 suggests that we obtain similar results when using the mixed indicator. Column 5 shows that the estimate almost doubles when the latent variable approach is used for data aggregation. This additional increase in the regression coefficient supports our econometric argument because the latent variable approach produces the greatest distortions for single-party regimes and established democracies (see Figure 6).[47]

[45] Equation (27) also suggests that the same logic applies when the measure of democracy is upward-biased for autocratic regimes:

$$d_s \;<\; \delta_s \;<\; \delta_t \;\leq\; d_t \quad \text{with} \quad s \in \{1, \ldots, \tau\} \text{ and } t \in \{\tau+1, \ldots, T\}.$$

[46] Appendix Table C.5 illustrates that the increase in the size of the effect is even more pronounced when using the alternative additive indicator (equal weights).

[47] Appendix Table C.6 shows that the results presented in Table 9 are robust to the inclusion of control variables (Panel A) and the use of the Difference GMM estimator (Panel B).

Table 9 Democracy and economic development — Alternative aggregation methods

	CSVMDI **(Baseline)**	**Additive** **(PCA)**	**Multiplicative** **(PCA)**	**Mixed** **(PCA)**	**Latent Variable**
	(1)	(2)	(3)	(4)	(5)
Democracy	0.0198*	0.0270	0.0262**	0.0278**	0.0379*
	(0.0107)	(0.0172)	(0.0117)	(0.0138)	(0.0200)
Log GDP^{pc} $(t-1)$	1.169***	1.167***	1.168***	1.168***	1.166***
	(0.0315)	(0.0314)	(0.0314)	(0.0314)	(0.0315)
Log GDP^{pc} $(t-2)$	-0.189***	-0.189***	-0.189***	-0.189***	-0.188***
	(0.0269)	(0.0269)	(0.0269)	(0.0269)	(0.0270)
Log GDP^{pc} $(t-3)$	0.0428	0.0424	0.0429	0.0426	0.0424
	(0.0303)	(0.0302)	(0.0303)	(0.0302)	(0.0302)
Log GDP^{pc} $(t-4)$	-0.0577***	-0.0570***	-0.0571***	-0.0570***	-0.0566***
	(0.0141)	(0.0140)	(0.0141)	(0.0140)	(0.0140)
Observations	7,606	7,606	7,606	7,606	7,606
Countries	170	170	170	170	170
CD F-Stat.	844.80	531.18	849.53	729.79	380.92
SY 10% IV size	16.38	16.38	16.38	16.38	16.38
AR p-value	0.06	0.12	0.02	0.04	0.05
SW p-value	0.04	0.10	0.02	0.03	0.05
OP F Stat.	111.72	90.22	115.47	103.68	89.26
OP τ =5%	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.002	-0.009	0.006	0.002	-0.003
UCI (upper)	0.038	0.063	0.047	0.053	0.079
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000
CSVMDI Ratio	–	1.364	1.323	1.404	1.914
R-Squared	0.96	0.96	0.96	0.96	0.96

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects and year fixed effects. The measures of democracy only differ with regard to the aggregation technique. The bottom part of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP $\tau = 5\%$. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**p < .05$, $***p < .01$.

Taken together, the empirical evidence presented in this section is suggestive of two important conclusions. First, the 2SLS estimator cannot compensate for the measurement errors that are produced by conventional aggregation techniques, and second, the economic consequences of a democratic transition are overestimated when conventional aggregation methods are used to create measures of democracy.

5 Conclusion

The question of how political institutions and economic development interact is arguably one of the most contentious in political economy. This study provides novel insights in this relationship and suggests that democracy and economic prosperity are more closely related than presumed so far.

The most important contribution of this paper is a new method for data aggregation.

This methodological innovation is a real step forward since we show that conventional aggregation techniques rest on ad hoc assumptions and thus give rise to subprime measures of democracy. The presented algorithm is based on a machine learning techniques for pattern recognition—known as Support Vector Machine (SVM)—and has four valuable features: (i) it does not require any assumption about the functional relationship between the constituent parts as given by the raw data and the resulting overall degree of democratization, (ii) it computes continuous and dichotomous indicators, (iii) it is flexible with regard to how the term "democracy" is conceptualized, and (iv) it provides indication of measurement uncertainty.

We especially elaborate on why there is a need for sophisticated aggregation methods. We show, for example, that the 2SLS estimator fails to compensate for the systematic malfunctions that are produced by conventional aggregation techniques. We also demonstrate that because of this failure, the effect of democracy on economic development is overestimated and explain why political economists should take uncertainty in measures of democracy into account.

Most studies on the socioeconomic consequences of democratization are based on conventional aggregation techniques and thus do not allow checking whether their results are robust to measurement uncertainty. We therefore suggest to replicate previous studies and to investigate whether controversial findings can be attributed to the shortcomings of conventional aggregation methods or to the uncertainty associated with the employed measures of democracy. The SVM algorithm can also help to answer open questions about the relationship between democracy and economic development. For example, one could compute separate indicators for the various dimensions of democracy and analyze whether some aspects are more conducive to economic development than others.

References

Abe, S. (2005). *Support vector machines for pattern classification*, volume 53. Springer. 4, 5, 6

Acemoglu, D. (2008). Oligarchic versus democratic societies. *Journal of the European Economic Association*, 6(1):1–44. 2

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and Democracy. *American Economic Review*, 98(3):808–42. 2

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2009). Reevaluating the modernization hypothesis. *Journal of Monetary Economics*, 56(8):1043–1058. 2

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2014). Democracy Does Cause Growth. NBER Working Papers 20004, National Bureau of Economic Research, Inc. 2, 3, 8, 13, 24, 25, 28, 45, 46

Acemoglu, D. and Robinson, J. A. (2006). Economic origins of democracy and dictatorship. 30

African Election Database (2017). *A database of election results in Sub-Saharan Africa.*

9, 10

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. 2

Aidt, T. S. and Jensen, P. S. (2014). Workers of the world, unite! franchise extensions and the threat of revolution in europe, 1820–1938. *European Economic Review*, 72:52–75. 25

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. 2

Aisen, A. and Veiga, F. J. (2013). How does political instability affect economic growth? *European Journal of Political Economy*, 29:151–167. 2

Alvarez, M., Cheibub, J. A., Limongi, F., and Przeworski, A. (1996). Classifying political regimes. *Studies in Comparative International Development*, 31(2):3–36. 12, 30

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. 26, 28

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*, volume 1. Princeton university press Princeton. 35

Ansell, B. W. (2010). *From the ballot to the blackboard: The redistributive political economy of education*. Cambridge University Press. 3

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):pp. 277–297. 30

Barro, R. J. (1999). Determinants of democracy. *Journal of Political economy*, 107(S6):S158–S183. 2

Beck, R. B. (2013). *The History of South Africa*. ABC-CLIO. 17, 18

Becker, J. (2004). Lessons from russia: A neo-authoritarian media system. *European Journal of Communication*, 19(2):139–163. 34

Bennett, K. P. and Campbell, C. (2000). Support vector machines: hype or hallelujah? *ACM SIGKDD Explorations Newsletter*, 2(2):1–13. 4

Besley, T. and Coate, S. (1998). Sources of inefficiency in a representative democracy: a dynamic analysis. *American Economic Review*, pages 139–156. 2

Besley, T. and Kudamatsu, M. (2006). Health and democracy. *The American economic review*, 96(2):313–318. 3

Besley, T., Persson, T., and Sturm, D. M. (2010). Political competition, policy and growth: theory and evidence from the us. *The Review of Economic Studies*, 77(4):1329–1352. 3

Bluhm, R., Gassebner, M., Langlotz, S., and Schaudt, P. (2016). Fueling Conflict? (De)Escalation and Bilateral Aid. Technical report. 27, 57, 58

Bogaards, M. (2010). Measures of democratization: From degree to type to war. *Political Research Quarterly*, 63(2):475–488. 46

Boix, C., Miller, M., and Rosato, S. (2013). A complete data set of political regimes, 1800–2007. *Comparative Political Studies*, 46(12):1523–1554. 13, 45, 46

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review*, pages 370–390. 9

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*, pages 144–152. ACM Press. 5, 6

Breiman, L. et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231. 4

Brückner, M. and Ciccone, A. (2011). Rain and the democratic window of opportunity. *Econometrica*, 79(3):923–947. 2, 25

Brückner, M., Ciccone, A., and Tesei, A. (2012). Oil price shocks, income, and democracy. *Review of Economics and Statistics*, 94(2):389–399. 2

Burke, P. J. and Leigh, A. (2010). Do output contractions trigger democratic change? *American Economic Journal: Macroeconomics*, 2(4):124–157. 2

Carr, A. (2017). *Psephos: Adam Carr's Electoral Archive*. Melbourne (AU). 9, 10

Cascio, E. U. and Washington, E. (2013). Valuing the vote: The redistribution of voting rights and state funds following the voting rights act of 1965. *The Quarterly Journal of Economics*, 129(1):379–433. 3

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

Chaney, E. (2013). Revolt on the nile: Economic shocks, religion, and political power. *Econometrica*, 81(5):2033–2053. 2

Chang, P. Y. (2008). Unintended consequences of repression: Alliance formation in south korea's democracy movement (1970-1979). *Social Forces*, 87(2):651–677. 18

Cheibub, J. A., Gandhi, J., and Vreeland, J. R. (2010). Democracy and dictatorship revisited. *Public Choice*, 143(1-2):67–101. 10, 13, 14, 30, 45, 46

Collier, D. and Adcock, R. (1999). Democracy and dichotomies: A pragmatic approach to choices about concepts. *Annual Review of Political Science*, 2:537–565. 30

Collier, D. and Levitsky, S. (1997). Democracy with adjectives: Conceptual innovation in comparative research. *World politics*, 49(03):430–451. 9

Collier, S. and Sater, W. F. (2004). *A history of Chile, 1808-2002*, volume 82. Cambridge University Press. 16

Conley, T., Hansen, C., and Rossi, P. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272. 26, 27, 29, 33, 36, 57, 58, 59, 60

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(02):247–267. 45

Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Altman, D., Bernhard, M., Fish, M. S., Glynn, A., Hicken, A., et al. (2016). V-dem [country-year/country-date] dataset v6.2. 10, 14, 34

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297. 5

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

Haven (US). 9

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. 3

Decancq, K. and Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, 32(1):7–34. 8

Devi, K. S. (2014). Myanmar under the military rule 1962-1988. *International Research Journal of Social Sciences*, 3(10):46–50. 16

Diamond, L., Linz, J., and Lipset, S. M. (1990). Politics in developing countries. *Comparing Experiences with Democracy, Boulder/London*. 10

Doucouliagos, H. and Ulubaşoğlu, M. A. (2008). Democracy and economic growth: a meta-analysis. *American Journal of Political Science*, 52(1):61–83. 2, 30

Dreher, A. (2006). Does globalization affect growth? evidence from a new index of globalization. *Applied Economics*, 38(10):1091–1110. 19

Durlauf, S. N., Johnson, P. A., and Temple, Jonathan R. W. (2005). Chapter 8 growth econometrics. In Philippe Aghion and Steven N. Durlauf, editor, *Handbook of Economic Growth*, volume Volume 1, Part A, pages 555–677. Elsevier. 25, 28

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *The American Economic Review*, 105(10):3150–3182. 26, 27, 29, 33, 36, 57, 58, 59, 60

Franck, R. (2016). The political consequences of income shocks: Explaining the consolidation of democracy in france. *Review of Economics and Statistics*, 98(1):57–82. 2

Freedom House (2015). *Freedom in the World database*. Washington D.C.(US). 14, 15, 45

Fujiwara, T. (2015). Voting technology, political responsiveness, and infant health: Evidence from brazil. *Econometrica*, 83(2):423–464. 3

Geddes, B., Wright, J., and Frantz, E. (2014). Autocratic breakdown and regime transitions: A new data set. *Perspectives on Politics*, 12(2):313–331. 14

Giuliano, P., Mishra, P., and Spilimbergo, A. (2013). Democracy and reforms: evidence from a new dataset. *American Economic Journal: Macroeconomics*, 5(4):179–204. 3, 25

Gleditsch, K. S. and Ward, M. D. (2006). Diffusion and the international context of democratization. *International organization*, 60(4):911–933. 25, 28

Goertz, G. (2006). *Social science concepts: A user's guide*. Princeton University Press. 8

Grosjean, P. and Senik, C. (2011). Democracy, market liberalization, and political preferences. *The Review of Economics and Statistics*, 93(1):365–381. 3

Gründler, K. and Krieger, T. (2016). Democracy and growth: Evidence of a machine learning indicator. *European Journal of Political Economy*, forthcoming. 2, 3, 45

Guha-Sapir, D., Below, R., and Hoyois, P. (2015). Em-dat: International disaster database. *Catholic University of Louvain: Brussels, Belgium*. 27, 57, 58

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. 2

Guttman, L. (1994). *Louis Guttman on theory and methodology: Selected writings*. Dartmouth Publishing Group. 9

Hadenius, A. (1992). *Democracy and development*. Cambridge University Press. 9

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. 17

Han, S.-j. (1974). *The failure of democracy in South Korea*, volume 35. Univ of California Press. 18

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054. 29

Harding, R. and Stasavage, D. (2013). What democracy does (and doesn't do) for basic services: School fees, school inputs, and african elections. *The Journal of Politics*, 76(1):229–245. 3

Huntington, S. P. (1993). *The third wave: Democratization in the late twentieth century*, volume 4. University of Oklahoma press. 25, 28

IDEA (2017). *Voter Turnout Database*. International Institute for Democracy and Electoral Assistance, Stockholm (SE). 9

IFES (2017). Election database. *International Foundation for Electoral Systems*. 9, 10

IPU (2017). *Parline databse on national parliaments*. Inter-Parliamentary Union, Le Grand-Saconnex (CH). 9, 10

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126. 26, 29, 33, 36, 57, 58, 59, 60

Kudamatsu, M. (2012). Has democratization reduced infant mortality in sub-saharan africa? evidence from micro data. *Journal of the European Economic Association*, 10(6):1294–1317. 3

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. 10

Lipset, S. M. (1959). Some social requisites of democracy. *American Political Science Review*, 53(1):69–105. 2

Lundström, S. (2005). The effect of democracy on different categories of economic freedom. *European Journal of Political Economy*, 21(4):967–980. 3

Madsen, J. B. and Murtin, F. (2017). British economic growth since 1270: the role of education. *Journal of Economic Growth*, 22(3):229–272. 2

Madsen, J. B., Raschky, P. A., and Skalil, A. (2015). Does democracy drive income in the world, 1500–2000? *European Economic Review*, 78:175—-195. 2

Márquez, X. (2016). A quick method for extending the unified democracy scores. 23

Marshall, M., Gurr, T. R., and Jaggers, K. (2016). *Polity IV project. Political Regime Characteristics and Transitions, 1800-2015*. Center of Systematic Peace. 8, 13, 14, 15, 45, 47, 50

Mattera, D. and Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In *Advances in kernel methods*, pages 211–241. MIT Press. 11

Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical transactions of the Royal Society, London A*, 209:415–446. 6

Munck, G. L. (2009). *Measuring democracy: A bridge between scholarship and politics*. JHU Press. 45

Munck, G. L. and Verkuilen, J. (2002). Conceptualizing and meassuring democracy: Evaluating alternative indices. *Comparative Political Studies*, 35(1):5–34. 3, 8, 9, 11, 45

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

Nohlen, D. (2005). *Elections in the Americas: A Data Handbook: Volume 2 South America*, volume 2. OUP Oxford. 9, 10, 16

Nohlen, D., Grotz, F., and Hartmann, C. (2001). *Elections in Asia and the Pacific: A Data Handbook*. Oxford University Press. 9, 10, 16, 18, 20

Nohlen, D. and Stöver, P. (2010). *Elections in Europe: A data handbook*. Nomos Verlag-Ges. 9, 10, 17

Nohlen, D., Thibaut, B., and Krennerich, M. (1999). *Elections in Africa: A Data Handbook: A Data Handbook*. OUP Oxford. 9, 10, 17, 18

Oates, S. (2007). The neo-soviet model of the media. *Europe-Asia Studies*, 59(8):1279–1297. 34

Olea, J. L. M. and Pflüger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369. 26

Olson, M. (1993). Dictatorship, democracy, and development. *American political science review*, 87(3):567–576. 13

O'Donnell, G. A. (2001). Democracy, law, and comparative politics. *Studies in Comparative International Development*, 36(1):7–36. 9

Papaioannou, E. and Siourounis, G. (2008). Democratisation and Growth. *Economic Journal*, 118(532):1520–1551. 2, 24

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. 8, 10, 13, 14, 15, 23, 45, 47, 50, 55

Persson, T. and Tabellini, G. (1994). Is Inequality Harmful for Growth? *American Economic Review*, 84(3):600–621. 2

Persson, T. and Tabellini, G. (2009). Democratic Capital: The Nexus of Political and Economic Change. *American Economic Journal: Macroeconomics*, 1(2):88–126. 25

Persson, T. and Tabellini, G. E. (2005). *The economic effects of constitutions*. MIT press. 14

Persson, T. and Tabellini, G. E. (2006). Democracy and Development: The Devil in the Details. *American Economic Review*, 96(2):319–324. 14

Plattner, M. F. (2015). Is democracy in decline? *Journal of Democracy*, 26(1):5–10. 2

Przeworski, A. (2000). *Democracy and development: Political institutions and well-being in the world, 1950-1990*, volume 3. Cambridge University Press. 2

Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond.* Wiley Online Library. 11

Rieffel, L. (2013). Myanmar on the move: an overview of recent developments. *Journal of Current Southeast Asian Affairs*, 31(4):31–49. 16

Rode, M. and Bjørnskov, C. (2016). Regime types and regime change: A new dataset. Technical report, Mimeo, Universidad de Navarra/Aarhus University. 14

Rode, M. and Gwartney, J. D. (2012). Does democratization facilitate economic liberalization? *European Journal of Political Economy*, 28(4):607–619. 3

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489. 28, 29

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons. 28, 29

Saint-Paul, G. and Verdier, T. (1993). Education, democracy and growth. *Journal of Development Economics*, 42(2):399–407. 2

Sakwa, R. (1990). *Gorbachev and his Reforms, 1985-1990.* Prentice Hall. 17

Sakwa, R. (2005). *The rise and fall of the Soviet Union.* Routledge. 17

Sakwa, R. (2014). *Putin Redux: Power and Contradiction in Contemporary Russia.* Routledge. 17

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

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319. 6

Schultz, J. (1998). *Reviving the fourth estate: Democracy, accountability and the media.* Cambridge University Press. 33

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy.* George Allen and Unwin. 32

Skaaning, S.-E., Gerring, J., and Bartusevičius, H. (2015). A lexical index of electoral democracy. *Comparative Political Studies*, 48(12):1491–1525. 13

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222. 4, 7

Steinwart, I. and Christmann, A. (2008). *Support vector machines.* Springer, New York (US). 4, 5, 6

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York. 26, 29, 33, 36, 57, 58, 59, 60

Stock, J. H. and Wright, J. H. (2000). Gmm with weak identification. *Econometrica*, 68(5):1055–1096. 26, 28

Tavares, J. and Wacziarg, R. (2001). How democracy affects growth. *European Economic Review*, 43(8):1341–1378. 3

Teorell, J., Coppedge, M., Skaaning, S.-E., and Lindberg, S. I. (2016). Measuring electoral democracy with v-dem data: Introducing a new polyarchy index. *V-Dem Working Paper*, 25. 8, 10, 13, 14, 15, 22, 24, 46, 47, 50, 55

Treier, S. and Jackman, S. (2008). Democracy as a latent variable. *American Journal of Political Science*, 52(1):201–217. 45

Vanhanen, T. (2000). A new dataset for measuring democracy, 1810-1998. *Journal of Peace Research*, 37(2):251–265. 8, 13, 45, 47, 50

Vapnik, V. (1995). The nature of statical learning theory. 4, 6, 7

Vapnik, V. N. (1998). *Statistical learning theory*, volume 1. Wiley New York. 4, 6, 7

Wike, R., Simmons, K., Stokes, B., and Fetterolf, J. (2017). Globally, broad support for representative and direct democracy. but many also endorse nondemocratic alternatives. pew research center. 2

World Bank (2017). *World Development Indicators Database (WDI).* The World Bank. 27, 57, 58

Appendix for Online Publication

A Supplementary comments

A.1 Supplementary notes to Section 3

In this section, we use three well-known measures of democracy to illustrate why indicators generated via conventional aggregation methods have often been criticized. For a more comprehensive literature review, see Boix et al. (2013), Cheibub et al. (2010), Munck and Verkuilen (2002), Munck (2009), Coppedge et al. (2011), and Gründler and Krieger (2016).

A multiplicative measure of democracy

Vanhanen (2000) assumes that the levels of electoral participation and political competition determine the degree of democratization. Participation is operationalized via the voter turnout rate and competition through the share of votes received by the leading party. The aggregation rule is multiplicative and the two characteristics are assumed to be equally important.

The literature emphasizes three major deficiencies of this approach: First, as noted by Vanhanen (2000), the weighting scheme is arbitrary. Second, it is presumed that a regime is fully democratic only if voter turnout is 100 percent. Finally, measurement uncertainty that arises from imprecise election statistics cannot be addressed, since the aggregation rule does not allow for the calculation of confidence intervals (Pemstein et al., 2010).

An additive measure of democracy

The Polity index of Marshall et al. (2016) is an ordinal measure that ranges from -10 (most autocratic) to $+10$ (most democratic) and comprises five characteristics determined by experts.[48] This measure lacks sophistication for a number of reasons: First, no theoretical justification is provided for the assumption that the characteristics affect the degree of democratization independently of one another or for why they are equally important. Second, the equal weighting leads to double counting since some of the characteristics are redundant. Finally, measurement uncertainties are not indicated via confidence intervals (Cheibub et al., 2010, Munck and Verkuilen, 2002, Treier and Jackman, 2008).

A mixed measure of democracy

Acemoglu et al. (2014) combine the information of additive and multiplicative measures. A regime is labeled "democratic" if Polity is positive and the additive index of Freedom House (2015) [F-House] indicates "free" or "partially free". The multiplicative measures

[48] The characteristics are equally weighted and referred to as: (i) competitiveness of political participation, (ii) regulation of political participation, (iii) openness of executive recruitment, (iv) competitiveness of executive recruitment, and (v) executive constraints. For details, see Marshall et al. (2016).

of Boix et al. (2013) and Cheibub et al. (2010) are used as additional decision criteria if Polity or F-House is not available.[49]

The approach of Acemoglu et al. (2014) has some weak points: First, a Polity score greater than 0 is assumed to be a necessary condition for the existence of democracy institution. This assumption may be problematic since (i) Polity is an imprecise measure, and (ii) the critical value is arbitrarily chosen.[50] Second, the concept that is reflected by the index cannot be explicitly specified because democracy is defined differently by each of the four sub-indices.[51] Third, the aggregation technique produces spurious regime transitions that must be removed manually (Acemoglu et al., 2014).[52] Finally, measurement uncertainty cannot be indicated.

A.2 Supplementary notes to Section 3.5

The analysis presented in Section 3.5 suggests that conventional aggregation methods produce upward-biased indicators for certain autocracies and downward-biased indices for established democracies. The consequence is an underestimation of institutional transitions. One may argue that this problem only arises because of a strategic selection of the characteristics. This section presents two supplementary analyses to allay this concern.

Alternative raw data

Teorell et al. (2016) use a concept of democratic institutions that requires the existence of electoral participation, political competition, and political liberties. The authors refer to this concept as "electoral democracy" and operationalize their definition through five expert-based ratings, capturing information on: suffrage, the procedure to select government officials, electoral fairness, freedom of expression, and freedom of association.

Teorell et al. (2016) present three measures of electoral democracy. The first is an additive index. In Appendix Figure B.5, we present the cases of Myanmar and Russia to illustrate that the additive indicator is upward-biased for certain autocracies. The second measure of electoral democracy is multiplicative. The example of New Zealand suggests that a downward distortion exists for established democracies (see Appendix Figure B.5). Finally, Teorell et al. (2016) make use of a mixed indicator, i.e. they calculate the average of the additive and the multiplicative measure of electoral democracy. Appendix Figure B.5 suggests that this index is upward-biased (downward-biased) for a number of autocratic regimes (democratic regimes).[53]

[49] There are two major reasons for missing information: (i) F-House is not available for the period from 1960 to 1972, and (ii) Polity is not available for small countries.

[50] For general concerns about the use of critical values, see Bogaards (2010) and Cheibub et al. (2010).

[51] Furthermore, the concept is not consistent over time due to changes in data availability. For instance, civil liberties are not included in the concept for some observations because the F-House was first published in 1973.

[52] Acemoglu et al. (2014) note explicitly that they must "*remove spurious transitions created when countries enter or leave the Freedom House, Polity, or our secondary sources' samples.*"

[53] For the sake of comparability, we use the SVM algorithm to produce our own measure of electoral democracy. Appendix Figure B.5 illustrates that this alternative indicator is not systematically biased

Taken together, these results show that our conclusion that the application of conventional aggregation methods leads to an underestimation of institutional transitions holds true if we use an alternative raw data set.

Alternative measures of democracy

In this subsection, we investigate whether standard measures of democracy are systematically biased at the upper and lower ends of the spectrum. We analyze four indicators: the additive index developed by Marshall et al. (2016) [Polity], the multiplicative index developed by Vanhanen (2000) [Polyarchy], the mixed index developed by Teorell et al. (2016) [TCSL], and the latent variable index developed by Pemstein et al. (2010) [UDS].

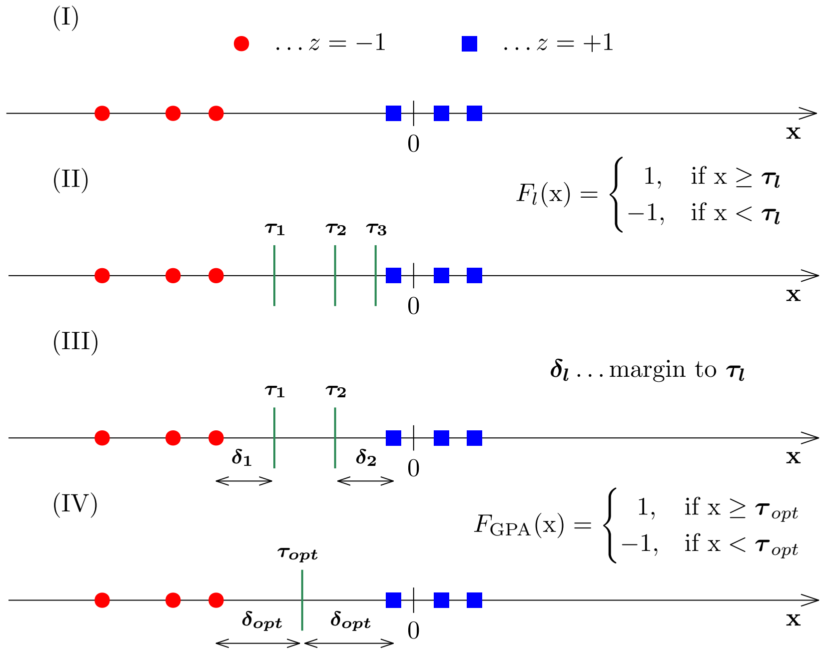
The upper section of Appendix Figure B.6 compares the density functions of alternative measures of democratic institutions. We observe that CSVMDI is bimodally distributed, with modes at the lowest and highest ends of the spectrum. Polity and TCSL have similar properties, however the location of the modes differs to some extent compared with CSVMDI. Furthermore, it turns out that the modes of UDS are located at intermediate degrees of democratization. We also observe that the distribution of Polyarchy is unimodal and skewed towards zero.

The lower section of Appendix Figure B.6 presents measures of democracy for New Zealand and Russia. These examples suggest that Polyarchy, TCSL and UDS (Polity, TCSL and UDS) are downward-biased (upward-biased) for established democracies (electoral autocracies).

at the upper or lower ends of the spectrum. This result strengthens the argument that the SVM algorithm outperforms conventional aggregation techniques.

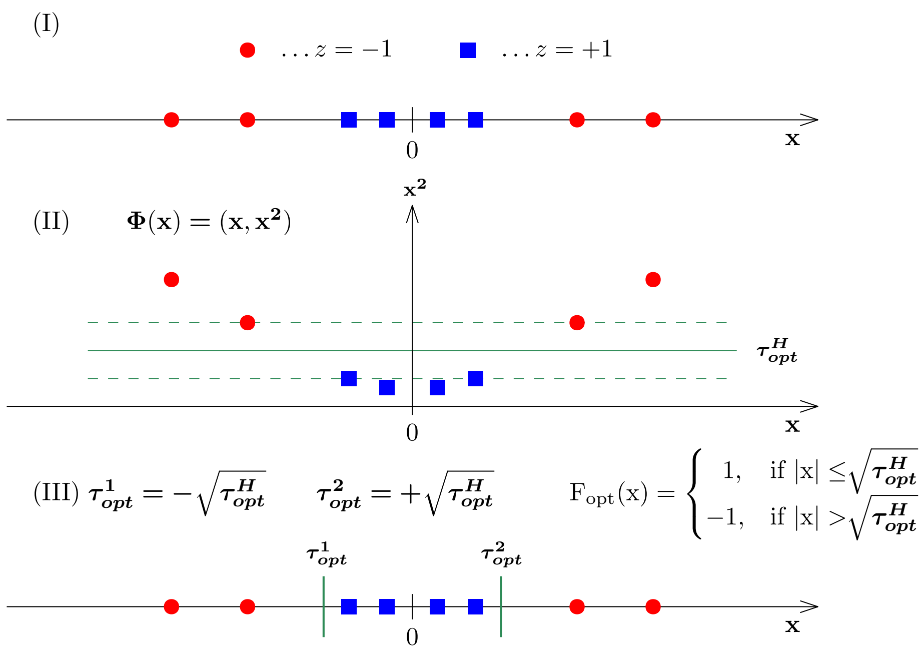
B Supplementary figures

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



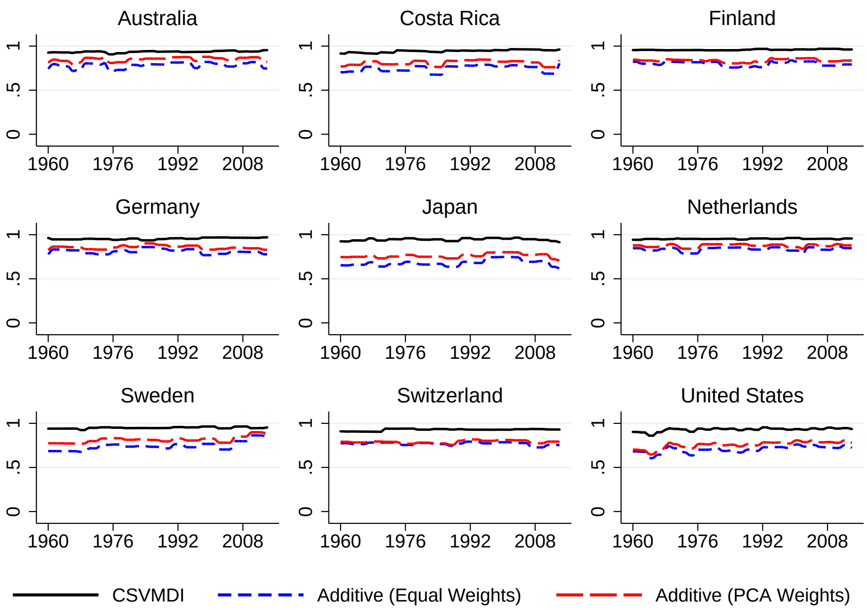
Notes: Graph I is a one-dimensional example in which GPA is applicable. Graph II shows that more than one hyperplane may separate the observations according to their labels. Graph III explains how the margin δ is calculated. Graph IV illustrates that GPA selects the hyperplane with the largest margin.

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



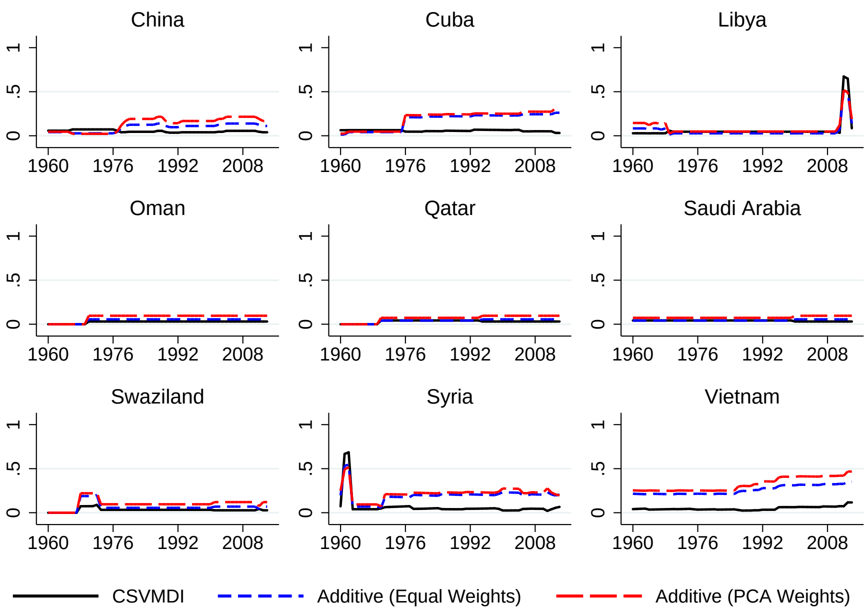
Notes: Graph I shows an example in which GPA is not applicable in $\mathcal{X} = \mathbb{R}$. In Graph II, a function $\Phi(x) = (x, x^2)$ is used to map the input data from $\mathcal{X} = \mathbb{R}$ onto a feature space $\mathcal{H} = \mathbb{R}^2$ and GPA computes a dividing hyperplane in $\mathcal{H}$.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

Figure B.3 CSVMDI vs. additive measures of democracy — Democratic regimes



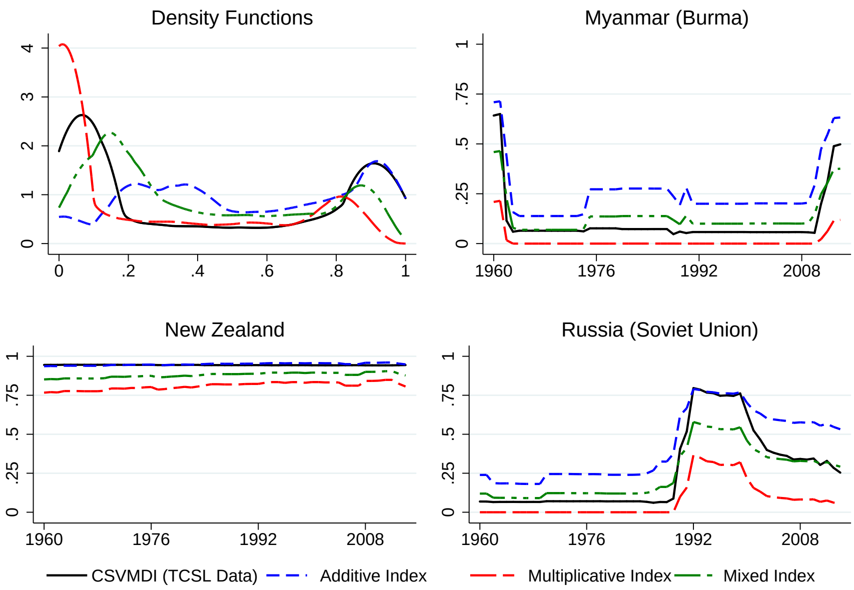
Notes: This figure compares CSVMDI with two additive measures of democracy. All indices are based on the same raw data, i.e. the differences in the degrees of democratization can be attributed to the particular aggregation method. The figure shows that additive aggregation methods produce downward-biased indices for established democracies.

Figure B.4 CSVMDI vs. additive measures of democracy — Autocratic regimes



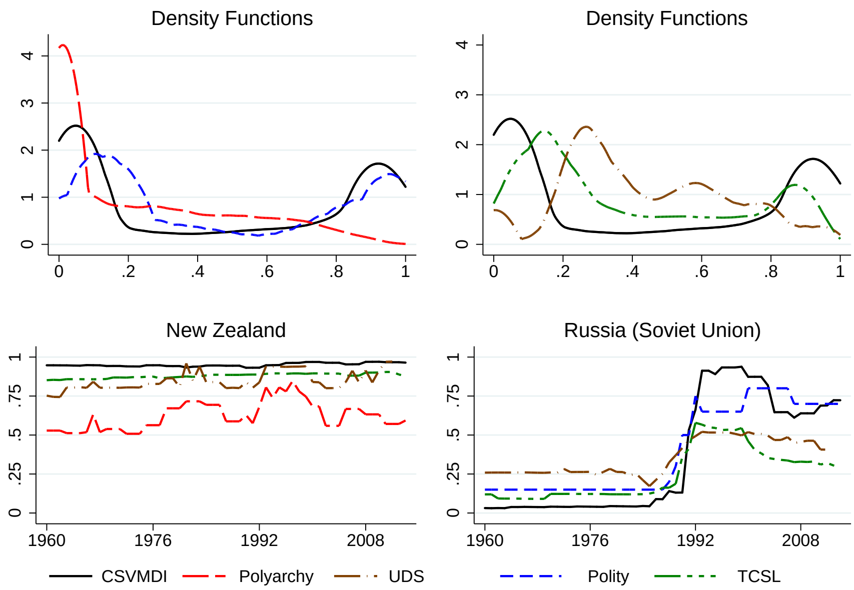
Notes: This figure compares CSVMDI with two additive measures of democracy. All indices are based on the same raw data, i.e. the differences in the degrees of democratization can be attributed to the particular aggregation method. The figure shows that additive aggregation methods produce upward-biased indices for some autocratic regimes.

Figure B.5 SVM algorithm vs. conventional aggregation methods — Alternative raw data



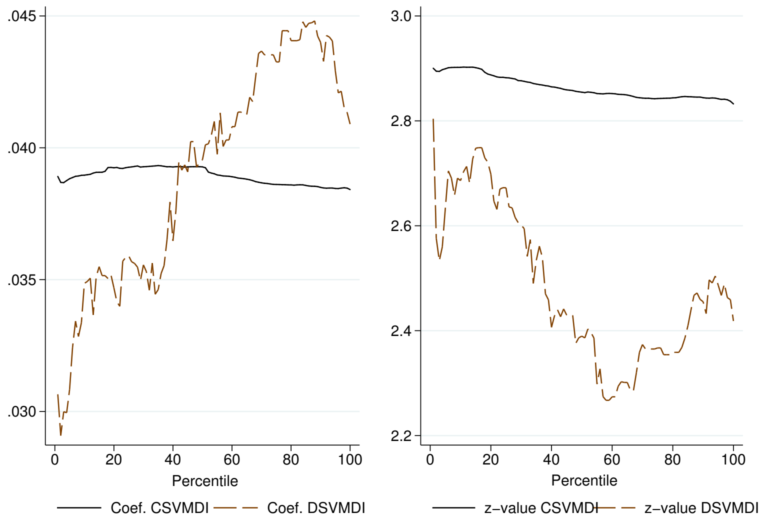
Notes: This figure compares the performance of different aggregation methods, using the conceptualization and operationalization proposed by Teorell et al. (2016). The figure allays concerns about strategic selection of the characteristics.

Figure B.6 Comparison of different measures of democracy



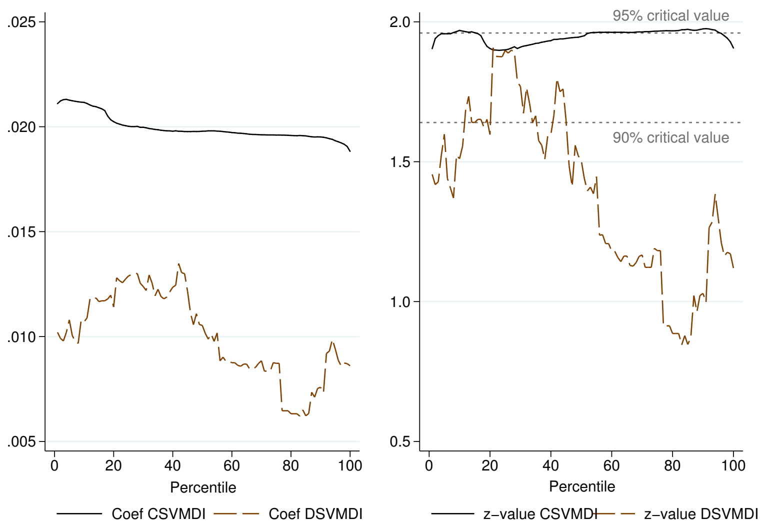
Notes: This figure compares CSVMDI with four measures of democracy. The alternative indicators are developed by Marshall et al. (2016) [Polity], Pemstein et al. (2010) [UDS], Teorell et al. (2016) [TCSL], and Vanhanen (2000) [Polyarchy].

Figure B.7 Consequences of measurement uncertainty — Augmented model



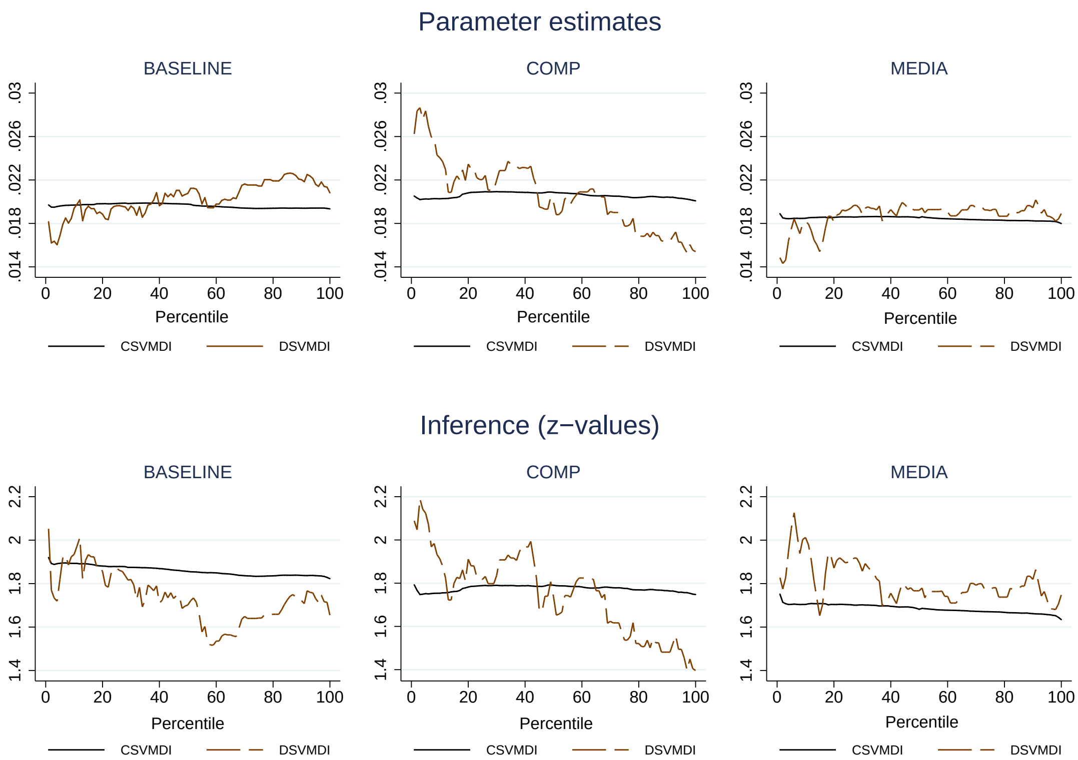
Notes: This figure investigates how the estimation results change when we take into account the measurement uncertainty in continuous and dichotomous measures of democracy. For this purpose, we use the distributions of the SVM indices (CDist, DDist) and estimate the regression model for every percentile of the distributions. The left graph reports the point estimates of the second-stage regressions, while the right graph presents the levels of statistical significance. All regressions control for the first four lags of the log of GDP per capita, the investment share ($t-1$), the inflation rate ($t-1$), the degree of trade openness ($t-1$), country fixed effects, and year fixed effects. The dependent variable is the log of GDP per capita. The regional degree of democratization in $t-1$ serves as the instrumental variable. The number of observations is 6832.

Figure B.8 Consequences of measurement uncertainty — Difference GMM



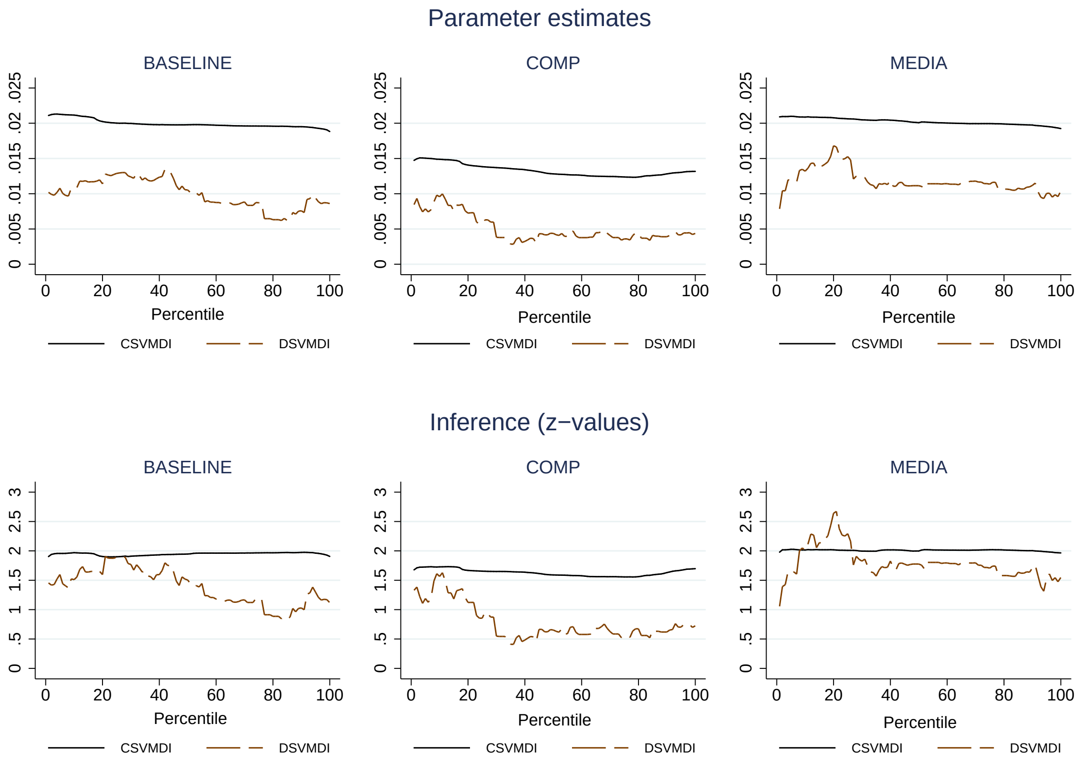
Notes: This figure investigates how the estimation results change when we take into account the measurement uncertainty in continuous and dichotomous measures of democracy. For this purpose, we use the distributions of the SVM indices (CDist, DDist) and estimate the regression model for every percentile of the distributions. The left graph reports the point estimates of the Difference GMM estimations, while the right graph presents the levels of statistical significance. All regressions control for the first four lags of the log of GDP per capita, country fixed effects, and year fixed effects. The dependent variable is the log of GDP per capita. The number of observations is 7436.

Figure B.9 Alternative concepts of democracy — Measurement Uncertainty (2SLS)



Notes: This figure investigates how the estimation results change when we take into account the measurement uncertainty in continuous and dichotomous measures of democracy. For this purpose, we use the distributions of the SVM indices and estimate the regression model for every percentile of the distributions. The baseline concept of democracy requires the existence of political competition, electoral participation, and political discourse. The concept "COMP" only requires the existence of political competition. The concept "MEDIA" urequires the existence of political competition, electoral participation, political discourse, and media independence. The upper graphs report the point estimates of the second-stage regressions, while the lower graphs present the levels of statistical significance. All regressions control for the first four lags of the log of GDP per capita, country fixed effects, and year fixed effects. The dependent variable is the log of GDP per capita. The regional degree of democratization in $t-1$ serves as the instrumental variable. The number of observations is 7606.

Figure B.10 Alternative concepts of democracy — Measurement Uncertainty (Difference GMM)



Notes: This figure investigates how the estimation results change when we take into account the measurement uncertainty in continuous and dichotomous measures of democracy. For this purpose, we use the distributions of the SVM indices and estimate the regression model for every percentile of the distributions. The baseline concept of democracy requires the existence of political competition, electoral participation, and political discourse. The concept "COMP" only requires the existence of political competition. The concept "MEDIA" requires the existence of political competition, electoral participation, political discourse, and media independence. The upper graphs report the point estimates of the Difference GMM regressions, while the lower graphs present the levels of statistical significance. All regressions control for the first four lags of the log of GDP per capita, country fixed effects, and year fixed effects. The dependent variable is the log of GDP per capita. The number of observations is 7436.

C Supplementary tables

Table C.1 Priming data — Selected country-years

Country	Observations	Years
Democratic regimes		
Australia	48	1964, 1966–67, 1969–70, 1972–2014
Austria	50	1961, 1963–65, 1967–2012
Barbados	2	1981–82
Belgium	52	1963–2014
Brazil	15	1998–2000, 2002–13
Canada	48	1972–81, 1984–85, 1988–2014
Chile	9	2005–08, 2010–14
Costa Rica	40	1975–2014
Cyprus	22	1989–2010
Czech Republic	24	1991–2014
Denmark	55	1960–2014
Estonia	23	1992–2014
Finland	33	1967–71, 1987–2014
France	45	1970–2014
Germany	44	1971–2014
Greece	23	1989–2011
Hungary	6	1994–97, 2004–05
Iceland	46	1967–2012
Ireland	32	1968, 1976–81, 1985, 1987, 1990–2012
Israel	1	1999
Italy	40	1960–61, 1963–76, 1983–92, 1994–07
Japan	8	1980–1982, 1984–86, 1988–89
Luxembourg	35	1964, 1967–71, 1977–98, 2000–01, 2006, 2008, 2010–12
Malta	27	1973–75, 1989–2012
Mauritius	1	2004
Netherlands	55	1960–2014
New Zealand	51	1963–2013
Norway	47	1968–2014
Poland	21	1991–2012
Portugal	27	1985–86, 1990-2014
Slovakia	2	2010, 2012
Slovenia	8	2003–04, 2009–14
Spain	35	1979–2013
St. Kitts and Nevis	1	2010
Sweden	55	1960–2014
Switzerland	45	1970–2014
United Kingdom	32	1961, 1964, 1975–1989, 2000–2014
United States	36	1969, 1976–77, 1980–81, 1984–2014
Uruguay	26	1989–2014
Autocratic regimes		
Afghanistan	28	1960–63, 1978–2001
Albania	23	1960–61, 1965, 1968, 1970, 1972–89
Algeria	14	1965–76, 1985, 1994
Angola	17	1975–91
Argentina	12	1966–1971, 1977–1982
Bahrain	13	1971, 1979–82, 1991–92, 1996–99
Benin	10	1973–78, 1984, 1986–87, 1989
Bhutan	47	1960–2007
Bolivia	7	1972–1977, 1980
Brazil	2	1968–69
Bulgaria	8	1980, 1983–89
Burkina Faso	1	1965
Myanmar (Burma)	39	1963–74, 1983–2008, 2010
Burundi	18	1967–81, 1988-89, 1991
Cambodia	16	1966–67, 1979–92
Cameroon	1	1988
Central African Rep.	16	1966–80, 1988
Chad	20	1962–63, 1970, 1972–74, 1976–89
Chile	14	1974–87
China	33	1960–79, 1985, 1989–97, 2000, 2013–14
Congo (Democratic Rep.)	17	1965–70, 1972–76, 1983–89
Congo (Rep.)	7	1968–72, 1978–79
Cuba	47	1960–2005, 2007
Dominican Republic	1	1960
Equatorial Guinea	17	1973–91
Eritrea	17	1994–96, 2001–14
Ethiopia	29	1960–71, 1973–86, 1988–1990
Gabon	1	1968
Germany (East)	25	1960–66, 1969, 1972–88
Ghana	1	1965
Greece	6	1968–1973
⋮	⋮	⋮

Table C.1 Priming data — Selected country-years (cont.)

Country	Observations	Years
⋮	⋮	⋮
Guatemala	3	1964–65, 1983
Guinea	15	1960, 1967–68, 1972–83
Guinea–Bissau	13	1977–1989
Haiti	21	1963–64, 1967–69, 1971–84, 1992–93
Indonesia	1	1965
Iran	12	1960–64, 1966–71, 1976
Iraq	40	1963–2002
Ivory Coast	2	1964–66
Jordan	18	1961–64, 1966–68, 1970, 1974–83
Korea (North)	55	1960–2014
Kuwait	12	1965–66, 1976–80, 1986–90
Laos	37	1976–88, 1991–2014
Lesotho	2	1970–71
Liberia	4	1980–83
Libya	42	1969–2010
Malawi	26	1964–66, 1968, 1970–87, 1989–92
Maldives	3	1965–67
Mali	7	1975–81, 1985
Mauritania	5	1979–83
Mongolia	7	1980, 1983–88
Morocco	9	1960–62, 1965–69, 1971
Mozambique	14	1976–89
Nepal	8	1960–65, 1967–68
Niger	12	1975–86
Nigeria	1	1966
Oman	21	1970–90
Pakistan	1	1980
Panama	3	1969–71
Paraguay	2	1960–61
Peru	3	1969, 1974–75
Philippines	5	1973–77
Portugal	9	1961–64, 1966–70
Qatar	45	1970–2014
Romania	6	1983–88
Russia (Soviet Union)	11	1962–69, 1984–86
Rwanda	5	1974–77, 1980
Saudi Arabia	55	1960–2014
Serbia (Yugoslavia)	20	1960–79
Somalia	19	1970–80, 1983–90
Spain	7	1960–66
Sudan	17	1960–63, 1969–70, 1989–97, 1999, 2001
Swaziland	11	1974–77, 1984, 1986–87, 1989–1992
Syria	30	1965, 1967–73, 1983–2002, 2011–12
Taiwan	9	1960–68
Togo	13	1968–79, 1985
Tunisia	1	1962
Turkmenistan	20	1992–2008, 2010–12
Uganda	8	1972–79
Uruguay	4	1976–79
Uzbekistan	16	1995–2006, 2008, 2010–12
Vietnam	14	1985–86, 1989–2000
Yemen	23	1960–70, 1976–1987

Notes: This table reports the country-years that are part of the priming data. The selected autocratic (democratic) observations belong to the lower (upper) decile of the index developed by Pemstein et al. (2010) or to the lower (upper) decile of the index developed by Teorell et al. (2016).

Table C.2 List of countries (by continent and region)

Continent	Region	List of countries
Africa	Central East Africa	Cameroon, Central African Republic, Chad, Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan
	North Africa	Algeria, Egypt, Libya, Morocco, Tunisia
	Southern Africa	Angola, Botswana, Burundi, Comoros, Democratic Rep. of Congo, Republic of Congo, Equatorial Guinea, Gabon, Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Rwanda, Sao Tome and Principe, Seychelles, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe
	Western Africa	Benin, Burkina Faso, Cape Verde, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, Togo
America	Caribbean	Antigua and Barbuda, Bahamas, Barbados, Cuba, Dominica, Dominican Republic, Grenada, Haiti, Jamaica, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago
	Central America	Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama
	North America	Canada, United States
	South America	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela
Asia and Oceania	Central Asia	Afghanistan, Armenia, Azerbaijan, Bhutan, Georgia, India, Iran, Kazakhstan, Kyrgyzstan, Maldives, Mongolia, Nepal, Pakistan, Sri Lanka, Tajikistan, Turkmenistan, Uzbekistan
	East and South-East Asia	Bangladesh, Myanmar (Burma), Cambodia, China, Japan, Korea (North), Korea (South), Laos, Taiwan, Thailand, Vietnam
	South Asia and Oceania	Australia, Brunei, East Timor, Fiji, Indonesia, Malaysia, New Zealand, Papua New Guinea, Philippines, Samoa, Singapore, Solomon Islands, Tonga, Vanuatu
	West Asia	Bahrain, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, Yemen
Europe	Balkan	Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Hungary, Kosovo, Macedonia, Montenegro, Romania, Serbia (Yugoslavia), Slovenia
	Central and North Europe	Austria, Belgium, Denmark, Finland, Germany, Iceland, Ireland, Luxembourg, Netherlands, Norway, Sweden, Switzerland, United Kingdom
	East Europe	Belarus, Czech Republic, East Germany, Estonia, Latvia, Moldova, Poland, Russia (Soviet Union), Slovakia, Ukraine
	South Europe	Cyprus, France, Greece, Italy, Malta, Portugal, Spain

Notes: This table lists all 186 countries for which SVM indices are available and categorizes these countries by continent and region.

Table C.3 Democracy and economic development — Domestic controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Second-stage estimates						
Democracy	0.0362***	0.0392***	0.0367***	0.0301**	0.0386***	0.0367***	0.0390***
	(0.0125)	(0.0138)	(0.0139)	(0.0122)	(0.0140)	(0.0137)	(0.0136)
Log GDP^{pc} $(t-1)$	1.182***	1.170***	1.166***	1.155***	1.171***	1.170***	1.170***
	(0.0343)	(0.0333)	(0.0348)	(0.0375)	(0.0339)	(0.0333)	(0.0332)
Log GDP^{pc} $(t-2)$	-0.190***	-0.190***	-0.182***	-0.165***	-0.184***	-0.191***	-0.190***
	(0.0317)	(0.0313)	(0.0320)	(0.0290)	(0.0297)	(0.0312)	(0.0313)
Log GDP^{pc} $(t-3)$	0.0351	0.0368	0.0379	0.0165	0.0292	0.0368	0.0367
	(0.0307)	(0.0301)	(0.0308)	(0.0324)	(0.0301)	(0.0298)	(0.0300)
Log GDP^{pc} $(t-4)$	-0.0595***	-0.0540***	-0.0611***	-0.0439***	-0.0515***	-0.0533***	-0.0534***
	(0.0138)	(0.0131)	(0.0131)	(0.0148)	(0.0134)	(0.0131)	(0.0127)
Investment $(t-1)$		0.0354*	0.0310	0.0293*	0.0303	0.0335*	0.0351*
		(0.0198)	(0.0202)	(0.0169)	(0.0199)	(0.0196)	(0.0196)
Inflation $(t-1)$		-0.0335**	-0.0306**	-0.0273**	-0.0322**	-0.0335**	-0.0337**
		(0.0159)	(0.0149)	(0.0133)	(0.0155)	(0.0156)	(0.0158)
Openness $(t-1)$		0.0181**	0.0179**	0.0105*	0.0193***	0.0189***	0.0184**
		(0.00709)	(0.00712)	(0.00598)	(0.00735)	(0.00690)	(0.00736)
Conflict $(t-1)$			-0.00284**				
			(0.00142)				
Human Capital $(t-1)$				0.0155*			
				(0.00889)			
Log Life Expect $(t-1)$					0.0357**		
					(0.0145)		
Earthquake $(t-1)$						0.00490*	
						(0.00297)	
Storm $(t-1)$						0.000418	
						(0.00170)	
Flood $(t-1)$						0.00444***	
						(0.00165)	
Drought $(t-1)$						0.00684***	
						(0.00230)	
Landslide $(t-1)$						0.00568**	
						(0.00273)	
Gov. Consumption $(t-1)$							-0.00667
							(0.0156)
	Panel B: First-stage estimates						
Democracy (Reg.)	0.6034***	0.5880***	0.5829***	0.5678***	0.5783***	0.5883***	0.5896***
	(0.0806)	(0.0839)	(0.0854)	(0.0903)	(0.0829)	(0.0837)	(0.0842)
Observations	6,832	6,832	6,631	5,972	6,729	6,822	6,832
Countries	169	169	169	141	169	169	169
CD F-Stat.	612.69	558.74	557.25	462.72	541.83	557.74	561.61
SY 10% IV size	16.38	16.38	16.38	16.38	16.38	16.38	16.38
AR p-value	0.00	0.00	0.00	0.01	0.00	0.00	0.00
SW p-value	0.00	0.00	0.00	0.01	0.00	0.00	0.00
OP F-Stat.	93.38	97.90	96.01	54.82	67.72	97.44	97.14
OP τ =5%	33.11	33.11	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.016	0.018	0.012	0.008	0.015	0.018	0.017
UCI (upper)	0.056	0.060	0.053	0.052	0.062	0.056	0.061
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Equal. (p-value)	–	0.823	0.971	0.619	0.861	0.973	0.836
R-Squared	0.96	0.96	0.96	0.96	0.96	0.96	0.96

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. Panel A reports second-stage estimates. Panel B presents first-stage estimates. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects and year fixed effects. The control variables are taken from Bluhm et al. (2016), Feenstra et al. (2015), Guha-Sapir et al. (2015) and World Bank (2017). The bottom section of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP τ = 5%. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). Equal (p-value) refers to the p-value of the Wald test for equality of parameter estimates. The base of comparison is the point estimates reported in Column (1). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**\,p < .05$, $***p < .01$.

Table C.4 Democracy and economic development — Regional controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Second-stage estimates						
Democracy	0.0356***	0.0391**	0.0347**	0.0303*	0.0386**	0.0375**	0.0390**
	(0.0123)	(0.0155)	(0.0151)	(0.0157)	(0.0154)	(0.0157)	(0.0156)
Log GDP^{pc} $(t-1)$	1.182***	1.181***	1.178***	1.163***	1.182***	1.180***	1.181***
	(0.0344)	(0.0346)	(0.0358)	(0.0382)	(0.0351)	(0.0345)	(0.0345)
Log GDP^{pc} $(t-2)$	-0.191***	-0.190***	-0.183***	-0.166***	-0.184***	-0.190***	-0.190***
	(0.0318)	(0.0319)	(0.0320)	(0.0294)	(0.0304)	(0.0318)	(0.0319)
Log GDP^{pc} $(t-3)$	0.0353	0.0351	0.0355	0.0154	0.0274	0.0352	0.0352
	(0.0307)	(0.0308)	(0.0316)	(0.0331)	(0.0309)	(0.0309)	(0.0308)
Log GDP^{pc} $(t-4)$	-0.0595***	-0.0591***	-0.0670***	-0.0458***	-0.0581***	-0.0591***	-0.0592***
	(0.0138)	(0.0137)	(0.0140)	(0.0145)	(0.0141)	(0.0137)	(0.0134)
Reg. Investment $(t-1)$		-0.00277	-0.00302	0.0200	-0.00854	-0.00462	-0.00252
		(0.0290)	(0.0279)	(0.0309)	(0.0298)	(0.0285)	(0.0289)
Reg. Inflation $(t-1)$		-0.0282	-0.00940	0.0331	-0.0213	-0.0318	-0.0289
		(0.0735)	(0.0740)	(0.0707)	(0.0742)	(0.0737)	(0.0743)
Reg. Openness $(t-1)$		0.00625	0.00372	0.000808	0.00664	0.00631	0.00603
		(0.00774)	(0.00759)	(0.00770)	(0.00785)	(0.00761)	(0.00778)
Reg. Conflict $(t-1)$			-0.00798***				
			(0.00287)				
Reg. Human Capital $(t-1)$				0.00753			
				(0.00890)			
Reg. Log Life Expect $(t-1)$					0.0138*		
					(0.00768)		
Reg. Earthquake $(t-1)$						0.00772	
						(0.00772)	
Reg. Storm $(t-1)$						0.00300	
						(0.00382)	
Reg. Flood $(t-1)$						0.00524	
						(0.00352)	
Reg. Drought $(t-1)$						-0.00940	
						(0.00583)	
Reg. Landslide $(t-1)$						0.00485	
						(0.00650)	
Reg. Gov. Consumption $(t-1)$							0.00265
							(0.0261)
	Panel B: First-stage estimates						
Democracy (Reg.)	0.6034***	0.5503***	0.5539***	0.5195***	0.5468***	0.5431***	0.5539***
	(0.0806)	(0.0919)	(0.0939)	(0.0990)	(0.0925)	(0.0926)	(0.0916)
Observations	6,822	6,822	6,631	5,962	6,719	6,822	6,822
Countries	169	169	169	141	169	169	169
CD F-Stat.	612.69	452.11	463.53	344.70	441.95	431.71	456.05
SY 10% IV size	16.38	16.38	16.38	16.38	16.38	16.38	16.38
AR p-value	0.00	0.00	0.01	0.03	0.01	0.01	0.01
SW p-value	0.00	0.00	0.01	0.03	0.01	0.01	0.00
OP F-Stat.	93.38	83.77	81.11	70.51	82.35	80.76	77.08
OP τ =5%	33.11	33.11	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.016	0.016	0.012	0.007	0.033	0.014	0.017
UCI (upper)	0.056	0.062	0.057	0.054	0.056	0.061	0.061
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Equal. (p-value)	–	0.852	0.919	0.707	0.876	0.933	0.858
R-Squared	0.96	0.96	0.96	0.96	0.96	0.96	0.96

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. Panel A reports second-stage estimates. Panel B presents first-stage estimates. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects and year fixed effects. The control variables are regional averages (jackknifed) and taken from Bluhm et al. (2016), Feenstra et al. (2015), Guha-Sapir et al. (2015) and World Bank (2017). The bottom section of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP $\tau = 5\%$. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). Equal (p-value) refers to the p-value of the Wald test for equality of parameter estimates. The base of comparison is the point estimates reported in Column (1). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**\,p < .05$, $***p < .01$.

Table C.5 Democracy and economic development — Alternative aggregation techniques

	CSVMDI (Baseline)	**Additive** (Equal)	**Multiplicative** (Equal)	**Mixed** (Equal)	**Latent Variable**
	(1)	(2)	(3)	(4)	(5)
Democracy	0.0198*	0.0354*	0.0307**	0.0334**	0.0379*
	(0.0107)	(0.0183)	(0.0131)	(0.0152)	(0.0200)
Log GDP^{pc} $(t-1)$	1.169***	1.167***	1.168***	1.167***	1.166***
	(0.0315)	(0.0316)	(0.0315)	(0.0315)	(0.0315)
Log GDP^{pc} $(t-2)$	-0.189***	-0.188***	-0.189***	-0.188***	-0.188***
	(0.0269)	(0.0270)	(0.0269)	(0.0270)	(0.0270)
Log GDP^{pc} $(t-3)$	0.0428	0.0425	0.0429	0.0427	0.0424
	(0.0303)	(0.0303)	(0.0303)	(0.0303)	(0.0302)
Log GDP^{pc} $(t-4)$	-0.0577***	-0.0568***	-0.0571***	-0.0570***	-0.0566***
	(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0140)
Observations	7,606	7,606	7,606	7,606	7,606
Countries	170	170	170	170	170
CF F-Stat.	844.80	463.02	746.28	629.55	380.92
SY 10% IV size	16.38	16.38	16.38	16.38	16.38
AR p-value	0.06	0.05	0.01	0.02	0.05
SW p-value	0.04	0.04	0.01	0.02	0.05
OP F Stat.	111.72	88.89	112.98	101.52	89.26
OP τ =5%	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.002	0.002	0.005	0.007	-0.003
UCI (upper)	0.038	0.068	0.056	0.060	0.079
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000
CSVMDI Ratio	–	1.788	1.551	1.687	1.914
R-Squared	0.96	0.96	0.96	0.96	0.96

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects and year fixed effects. The measures of democracy only differ with regard to the aggregation technique. The bottom part of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP $\tau = 5\%$. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**p < .05$, $***p < .01$.

Table C.6 Democracy and economic development — Alternative aggregation techniques

	CSVMDI (Baseline)	**Additive (PCA)**	**Multiplicative (PCA)**	**Mixed (PCA)**	**Latent Variable**
	(1)	(2)	(3)	(4)	(5)
	Panel A: Second-stage estimates				
Democracy	0.0392***	0.0530***	0.0473***	0.0509***	0.0636***
	(0.0138)	(0.0200)	(0.0151)	(0.0170)	(0.0241)
Log GDP^{pc} $(t-1)$	1.170***	1.168***	1.169***	1.168***	1.167***
	(0.0333)	(0.0339)	(0.0332)	(0.0335)	(0.0340)
Log GDP^{pc} $(t-2)$	-0.190***	-0.190***	-0.191***	-0.190***	-0.190***
	(0.0313)	(0.0313)	(0.0313)	(0.0313)	(0.0314)
Log GDP^{pc} $(t-3)$	0.0368	0.0358	0.0369	0.0364	0.0356
	(0.0301)	(0.0298)	(0.0301)	(0.0300)	(0.0297)
Log GDP^{pc} $(t-4)$	-0.0540***	-0.0529***	-0.0527***	-0.0527***	-0.0521***
	(0.0131)	(0.0129)	(0.0130)	(0.0130)	(0.0130)
Investment $(t-1)$	0.0354*	0.0357*	0.0367*	0.0361*	0.0356*
	(0.0198)	(0.0199)	(0.0197)	(0.0197)	(0.0202)
Openness $(t-1)$	0.0181**	0.0178***	0.0173**	0.0176**	0.0175***
	(0.00709)	(0.00685)	(0.00693)	(0.00689)	(0.00677)
Inflation $(t-1)$	-0.0335**	-0.0343**	-0.0346**	-0.0345**	-0.0340**
	(0.0159)	(0.0162)	(0.0161)	(0.0162)	(0.0166)
Observations	6,832	6,832	6,832	6,832	6,832
Countries	169	169	169	169	169
CD F-Stat.	558.74	418.60	594.22	545.09	310.27
SY 10% IV size	16.38	16.38	16.38	16.38	16.38
AR p-value	0.00	0.00	0.00	0.00	0.00
SW p-value	0.00	0.00	0.00	0.00	0.00
OP F Stat.	97.90	83.63	99.37	92.81	83.97
OP τ =5%	33.11	33.11	33.11	33.11	33.11
UCI (lower)	0.018	0.014	0.024	0.023	0.020
UCI (upper)	0.060	0.092	0.070	0.079	0.107
KP rk LM p-value	0.000	0.000	0.000	0.000	0.000
CSVMDI Ratio	–	1.352	1.207	1.298	1.622
R-Squared	0.96	0.96	0.96	0.96	0.96
	Panel B: Difference GMM estimates				
Democracy	0.0198*	0.0254**	0.0228**	0.0262**	0.0250**
	(0.0101)	(0.0107)	(0.0106)	(0.0111)	(0.0200)
Log GDP^{pc} $(t-1)$	1.418***	1.306***	1.371***	1.341***	1.305***
	(0.102)	(0.109)	(0.0993)	(0.104)	(0.103)
Log GDP^{pc} $(t-2)$	-0.446***	-0.295**	-0.394***	-0.346***	-0.301**
	(0.129)	(0.136)	(0.129)	(0.133)	(0.128)
Log GDP^{pc} $(t-3)$	0.0778**	0.0477	0.0707*	0.0577	0.0533
	(0.0363)	(0.0363)	(0.0365)	(0.0364)	(0.0349)
Log GDP^{pc} $(t-4)$	-0.0518**	-0.0475**	-0.0525**	-0.0496**	-0.0520**
	(0.0228)	(0.0221)	(0.0220)	(0.0220)	(0.0212)
Observations	7,436	7,436	7,436	7,436	7,436
Countries	170	170	170	170	170
Hansen p-value	0.072	0.038	0.041	0.035	0.027
AR(1) p-value	0.000	0.001	0.000	0.000	0.000
AR(2) p-value	0.201	0.799	0.327	0.546	0.724
Instruments	120	120	120	120	120
CSVMDI ratio	–	1.283	1.152	1.323	1.263

Notes: Using data from Feenstra et al. (2015), the dependent variable is the log of GDP per capita. The unit of observation is the country-year. Panel A presents second stage estimates and Panel B reports Difference GMM estimates. The standard errors are reported in parentheses, are clustered at the country level, and are robust to heteroscedasticity. The instrumental variable is the regional degree of democratization in year $t-1$ (jackknifed). All regressions include country fixed effects and year fixed effects. The measures of democracy only differ with regard to the aggregation technique. The bottom part of the table reports a number of test statistics to substantiate the strength and the validity of the instrumental variable. The Anderson-Rubin (AR) Wald test evaluates the null hypothesis $H_0 : \gamma = 0$, and is robust to weak instrumentation. The Stock-Wright (SW) test analyzes the same hypothesis, using a Lagrange Multiplier (LM) procedure. The Cragg-Donald (CD) F-statistic indicates the overall strength of the instrumental variable. The related critical value is denoted by SY 10% IV size and is taken from Stock and Yogo (2005). Since the CD F-statistic may be invalid in the presence of cluster-robust standard errors, the table also reports the Olea and Pflüger (OP) F-statistic, which allows for standard errors that are clustered, heteroscedastic, and serially correlated. The related critical value is denoted by OP $\tau = 5\%$. The KP rk LM p-value documents the p-value of the underidentification test proposed by Kleibergen and Paap (2006). UCI (lower) and UCI (upper) indicate the boundaries of the union of confidence interval (UCI) test introduced by Conley et al. (2012). The following notation is used to highlight coefficients that are significantly different from zero: $*p < .10$, $**\,p < .05$, $***p < .01$.