

# Is there a decline of democracy in the EU between 2004 and 2016? The relevance of data selection: a replication study of Smolka (2021) and comparison of democracy measures

Lukas Lemm

Accepted: 3 December 2021 / Published online: 20 December 2021  
© The Author(s) 2021

**Abstract** Is there a Decline of Democracy? Democracy measurement provides the basis for answering this question. However, there are different measurement tools based on different meanings of democracy that have been shown to vary in their concept validity. Therefore, it is relevant to examine whether the results of the different measurement tools converge or diverge with respect to a potential decline of democracy. Smolka (2021) finds a decline of democracy for new and old EU states based on standardized data from the Democracy Barometer. A re-analysis using the original data of the Democracy Barometer and the Democracy Matrix can hardly replicate these results. A comparison of further measurements shows that the instruments diverge rather than converge. I therefore conclude with some thoughts on overcoming the selection problem that arises in light of these contrasts.

**Keywords** Quality of democracy · Decline of democracy · Replication studies · Robustness checks · Measuring democracy · EU member states

---

Lukas Lemm, M. A. (✉)

Lehrstuhl für Vergleichende Politikwissenschaft und Systemlehre Institut für Politikwissenschaft und Soziologie, Julius-Maximilians-Universität Würzburg, Wittelsbacherplatz 1, 97074 Würzburg, Germany

E-Mail: [lukas.lemm@uni-wuerzburg.de](mailto:lukas.lemm@uni-wuerzburg.de)

## Gibt es zwischen 2004 und 2016 einen Rückgang der Demokratiequalität in der EU? Die Relevanz der Datenauswahl: eine Replikationsstudie von Smolka (2021) und ein Vergleich von Demokratiemessungen

**Zusammenfassung** Gibt es einen Verfall der Demokratie? Die Messung der Demokratie bietet die Grundlage für die Beantwortung dieser Frage. Allerdings gibt es verschiedene Messinstrumente, die auf unterschiedlichen Bedeutungen von Demokratie beruhen und sich in ihrer Konzeptvalidität unterscheiden. Daher ist es von Bedeutung zu untersuchen, ob die Ergebnisse der verschiedenen Messinstrumente in Bezug auf einen möglichen Rückgang der Demokratie konvergieren oder divergieren. Smolka (2021) stellt auf der Grundlage der standardisierten Daten des Demokratiebarometers einen Rückgang der Demokratie in den neuen und alten EU-Staaten fest. Eine Re-Analyse mit den Originaldaten des Demokratiebarometers und der Demokratiematrix kann diese Ergebnisse kaum replizieren. Ein Vergleich weiterer Messungen zeigt, dass die Instrumente eher divergieren als konvergieren. Ich schließe daher mit einigen Überlegungen zur Überwindung des Selektionsproblems, das sich angesichts dieser Kontraste aufdrängt.

**Schlüsselwörter** Qualität der Demokratie · Verfall der Demokratie · Replikationsstudien · Robustheitsprüfung · Messung der Demokratie · EU Mitgliedsstaaten

### 1 Introduction<sup>1</sup>

The fact that democracy is under attack is undisputed in political science and most societal debates. Less uniform, however, is the assessment of how far the consequences of these attacks go. The debate about a “crisis of democracy”<sup>2</sup> is multifaceted and the evaluations range from de-democratization processes within the democratic spectrum (Cianetti et al. 2018) to the identification of a new wave of autocratization (Lührmann and Lindberg 2019). However, there are also more optimistic positions that point to contrary developments like the liberalization of autocratic regimes and cases of democratization as well as the astonishing persistence of democratic regimes under unfavorable conditions (Levitsky and Way 2015, p. 56). As others before, a study by Smolka (2021) also deals with the measurement of democratic decline, analyzing developments among the EU member states. The starting point in this context is Article 2 of the Treaty on the EU, which sets out central normative values for the member states, but does not specify them in more detail, leaving a margin of interpretation open.

<sup>1</sup> Special thanks go to Oliver Schlenkrich, who provided some of the analyses, and to Hans-Joachim Lauth for his comments.

<sup>2</sup> While some still put a question mark on the debate title, others have already set it up so that it is followed by an exclamation mark. As the process is still underway, it is too early to decide on its outcome.

Here, democracy measurement as a sub-discipline of comparative political science steps in, which is tasked with mediating between theoretical concepts and empirical phenomena. However, there is also no agreement on what democracy is and how its quality should be assessed. Thus, the variety of different interpretations of the current state of democracy can be attributed to the conceptual differences of measurement instruments, which are also demonstrably of varying conceptual validity (Pickel et al. 2015; Munck 2016). The current expectation of a reversed wave is strongly linked to Huntington's (1991) metaphor of ebbs and flows,<sup>3</sup> although various studies have already shown that these waves—especially reversed waves—cannot be replicated with every measurement instrument and conceptual variation (Doorenspleet 2000; Møller and Skaaning 2013; Schmitter 2015, p. 33; Skaaning 2020). Thus, this research note builds on this preliminary works and attempts to answer the question of a potential decline of democracy—focusing on the EU member states—with a meta-analytical perspective. Do the empirical results of measurement instruments converge or diverge?

First, the results of Smolka's (2021) study, which uses standardized data from the *Democracy Barometer* (DB) is replicated by the data of the *Democracy Matrix* (DeMaX) and the original data from the DB. Then, the results of various other measurement tools are included to show the developments of the EU member states and to check whether there is indeed a meta-trend of decline of democracy. Finally, the results and their implications for research are discussed, and possible strategies for solving the selection problem of appropriate measurements are explored.

## 2 Replication of Smolka (2021): is there a decline of democracy among the EU member states?

We start by replicating the findings of Smolka (2021), who used the *Democracy Barometer* (DB) to analyze the development trends among the EU member states, with the data from the *Democracy Matrix* (DeMaX), which is a customized version of the *Varieties of Democracy Project* (V-Dem).<sup>4</sup> The guiding question is if these two measurement instruments agree or disagree regarding the EU development trends. Before we start, some preliminary words about these two measurement instruments are warranted. Both the DB and DeMaX share conceptual similarities concerning the three-dimensional structure, which is organized around the fundamental principles of democracy—freedom, equality, and control. Thus, both concepts allow the operationalization of Erdmann's (2011) approach to analyze the decline of democracy. And in this regard, they vary from other measurement concepts, which are organized

<sup>3</sup> To Way and Levitsky (2015) this adherence to a mental figure explains why the position of a reversed wave is so present in the debate (see exemplarily the title "Is the tide turning?" by Puddington 2008), even though it runs counter to empirical findings. Similarly, Mounk (2020) sees the paradigm of the end of history turned upside down and concludes that the prophecies of the pessimists replacing those of the optimists (see also Cianetti and Hanley 2021, pp. 67–69).

<sup>4</sup> Please contact the author for replication materials.

with a focus on concrete institutions on a lower level of abstraction and not around the principles of democracy on the most abstract level.

However, the conceptual trees (see appendix Fig. 3) show that both concepts differ when it comes to the explication of these three dimensions (see also Lauth 2010, p. 518): Whereas the DB subsumes institutions completely to certain dimensions, the DeMaX let the dimensions cut across the institutions. In other words: The DB measures the principles by specific institutions, whereas the DeMaX measures the institutions by the abstract principles. To illustrate this difference we take a look at the institution of the Rule of Law (RoL): While the DB subsumes it exclusively to the dimension of freedom, the DeMaX measure subparts of the RoL in each of the matrix fields of the institution Guarantee of Rights (GR) resulting in a three-dimensional measurement of the RoL, which comprises the independence of the judiciary (Freedom/GR), the equality before the law (Equality/GR), and the effective jurisprudence (Control/GR). Thus, the DB strives to measure freedom, but also assesses the equality before the law, which is one of the two sub-components of the RoL in the measurement scheme and clearly covers aspects of equality. Consequently, the DB disentangles the principles by institutions, which is partly at cost of conflation problems, while the DeMaX intertwines them.

There are several other noteworthy differences between both measurements, which can be only listed briefly: Whereas the DeMaX covers the whole regime continuum from autocracy to democracy, the DB captures the democratic spectrum which is preceded by a pre-selection of democracies—the so-called blueprint sample. In contrast to the DeMaX, which is based exclusively on V-Dem's expert ratings, the DB also integrates surveys into its measurement. The validity of these indicators is problematic, as it is not clear whether they actually measure what they are supposed to and, moreover, their cross-cultural comparability is questionable (Ariely and Davidov 2012; Knutsen and Wegmann 2016).<sup>5</sup> In addition, the DB also uses indicators that adopt the meaning of the quality of democracy in terms of results (see Diamond and Morlino 2004). This is problematic in that the “degree of association” does not directly capture the nature of institutions in terms of procedural quality. In other words, while a high degree of association is certainly desirable for a democracy, the procedural rules may still be weakly democratic.<sup>6</sup> The DeMaX refrains from including the quality of results with the exception of the context indicator on educational inequality preventing citizens from using their rights according to the concept of “low intensity citizenship” (Pinheiro 1999). Furthermore, the DB often

<sup>5</sup> There are some voices that suggest integrating improved subjective data collected by surveys into measurements of quality of democracy besides objective data, mostly expert ratings (Pickel et al. 2016). Fuchs and Roller (2018, p. 31) find a higher variance for the subjective than for the objective measurements of quality of democracy, which they trace back to different cultural contexts and historical traditions. And that is indeed problematic if we think of the DB indicator on the trust in the police, where low values could signal a low quality for democracy or only the fact that citizens are still suspicious due to the long shadow of autocratic rule, even though police reforms had already improved the quality of democracy.

<sup>6</sup> Even though the “degrees of association for economic interests and public interests” are offset against the “constitutional provisions guaranteeing freedom to associate”, the aggregation rule allows for compensation. This becomes clearer with another example: Even though elections are not held freely and fairly, a country can still achieve a strong score in the quality of democracy if voter turnout was high.

faces the accusation that some indicators in the sense of functional equivalents are not equally applicable to all institutional designs of democracies (Pickel et al. 2015, p. 511). For this reason, the DeMaX establishes the trade-off measurement as a third measurement level, which deals with these democracy profiles that are neutral with regard to the quality of democracy (Lauth and Schlenkrich 2018). For a more detailed discussion of the Democracy Barometer see the debate in this journal (Jäckle et al. 2012, 2013; Merkel et al. 2013). A critical evaluation of the validity of the DeMaX is still lacking.

In accordance with the Erdmann's (2011) approach, Smolka (2021) classifies a country as experiencing a decline of democracy (DoD) if it fulfills three criteria: 1) negative change of overall score **and** 2) negative change of control dimension **and** 3) negative change of the dimension freedom **or** equality.<sup>7</sup> Since it is relevant for the evaluation of a development dynamic to record the opposite changes, the mirror-image application of the rule should capture an improvement of democracy.<sup>8</sup> All cases that do not meet the criteria for either a decline or an improvement of democracy are grouped together in a residual category.<sup>9</sup>

When we apply the Erdmann-Smolka classification rule to the DeMaX data,<sup>10</sup> we find slightly more cases of decline of democracy (28) over all periods than Smolka (25) did. However, only eight cases are classified as declines of democracy by both measurements.<sup>11</sup> Furthermore, the DoDs take place in different periods: According to Smolka the majority of DoDs (14) appear between 2008 and 2012, whereas the DeMaX sees 16 cases of decline in the later period between 2012 and 2016. If we also include the cases of improvement of democracy (IoD) and relate them to the cases of DoD we note a divergent trend: Even though the democracy matrix states a much clearer trend (IoD/DoD = 21/3) than Smolka (14/5), both measurements agree with regard to a positive development of QdD in the first period. Then, in the second period, Smolka identifies a clear preponderance of DoDs over IoDs (4/14), whereas the ratio is reversed for DeMaX (15/9), even though the increasing number of DoDs is already becoming clear. In the third period, Smolka's negative trend weakens but remains (5/6). In the case of DeMaX, there is a true break in the QoD, since the

<sup>7</sup> Smolka made minor classification errors during calibration: Belgium also met the criteria for a decline of democracy in the period 2008–2012, as did Malta in the period 2012–2016. By contrast, Belgium, Estonia and Lithuania were incorrectly recorded as declines of democracy, although they did not show negative changes in the dimensions of freedom or equality, which is why they miss one of the criteria. The corrected figures can be found in Fig. 1.

<sup>8</sup> The improvement of democracy is given when the overall value **and** the value for control **and** the value for freedom **or** equality increases.

<sup>9</sup> This is not the same as stability or non-substantial change, as there can be either positive or negative changes as well, which is why they will only be referred to as insignificant.

<sup>10</sup> Unlike Smolka (2021, pp. 89–90), we use the original DeMaX data for the calculation and do not apply the suggested standardization, which takes the old EU member states between 1993 and 2003 as a baseline to rescale the values for the EU-28 between 2003 and 2016. Jäckle et al. (2012, pp. 114–115) already criticized that the min–max-standardization artificially generates variance and thus makes the countries' differences appear larger than they are.

<sup>11</sup> Austria, Portugal, Spain, Bulgaria and Hungary (each 2008–2012) as well as UK, Poland, Hungary and Malta (2012–2016).

16 DoDs are not opposed by any IoDs (0/16). Consequently, both measurements report deviating trends.

Interestingly, DoDs occur in roughly equal proportions among the old and new EU member states, even slightly more frequently among the old ones. In addition, the Smolka analysis reveals that IoDs are seldom among old EU member states, especially after 2008, whereas the DeMaX based analysis shows that IoDs among the old member states occur proportionally just as often as among the new ones. An evaluation of the split trends illustrates that for Smolka, DoDs outnumber IoDs in the old member states (IoD/DoD=8/14) and vice versa in the new member states (15/11), whereas according to DeMaX there is the described trend of an increase in QoD followed by a drop, which on the whole remains positive in terms of the ratio of DoDs and IoDs for the groups of the old (20/17) and new (16/11) member states. Thus, in Smolka's case, a possible decline of democracy is more localized in the old than in the new EU and starts in the phase of the financial crisis, whereas the DeMaX shows the decline being scattered across the EU and only starting in the third observed period.

To summarize the sharp contrasts of results between both measurement instruments, we can take a look at the total agreement regarding the three categorical classification: With the exception of Hungary and Poland they do not match any other country over all three periods and only a bit more than one third of compared country trends match (29 of 79 exclusively the missing cases).

Some of these results are caused by the fact that the classification rule does not take into account how large changes in quality of democracy are, which means that differences close to zero are interpreted as decline, although they should be better characterized as stagnation or stability. This is in line with suggestions in the literature (Waldner and Lust 2018, p. 95; Lührmann and Lindberg 2019, p. 1101), since not every negative change constitutes a decline of democracy. Thus, we recalculated the three categories after setting a minimum threshold for changes at 1% of the total scale span (see columns 1% rule in Fig. 1), which is  $\pm 1$  for the DB (0 to 100) and  $\pm 0.01$  for the DeMaX (0 to 1).<sup>12</sup> This should be perceived as a relatively lax threshold if we have in mind that the periods cover four years.

Thus, we still count 16 cases of decline of democracy using the Smolka data, but only nine for the DeMaX. The number of cases previously classified as improvements of democracy also decreased tremendously, which is why it tends to create the impression of stability instead of change in any direction. Interestingly—but somewhat expected since the residual category increased—the overall agreement between the measurements increased from 29 to 47 cases (59%) regarding the bivariate comparisons for the three categories. Relating improvements to declines of democracy once again, we see the same trends as above, even though they are reduced in sizes: The Smolka data indicate a positive trend (04–08: 7/3) followed by negative trend (08–12: 4/8) that weakens over time (12–16: 3/5), whereas the DeMaX reports a positive trend (04–08: 5/0) followed by stagnation (08–12: 2/2) and in the third period declines clearly exceed improvements (12–16: 0/7). After

<sup>12</sup> Of course, the threshold setting can be criticized at this point, as it is arbitrary rather than well-founded, but the fact that the results change so much at such a lax threshold clearly indicates that they are not robust.

Rule	Smolka 2021 (standardized Democracy Barometer)						Democracy Matrix (original data)						Democracy Barometer (original data)		
	Smolka rule			1% rule			Smolka rule			1% rule			Smolka rule		
Period	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16
Austria	1	-1	0	1	-1	0	1	-1	-1	0	0	0	0	0	0
Belgium	1	-1	0	1	-1	0	1	1	0	0	0	0	-1	1	1
Portugal	1	-1	0	0	0	0	-1	-1	0	0	0	0	0	1	1
France	1	-1	0	1	0	0	1	1	-1	0	0	0	0	1	-1
Greece	0	-1	0	0	-1	0	0	1	-1	0	0	0	0	-1	0
Sweden	1	0	1	0	0	1	1	-1	-1	0	0	0	0	1	1
Italy	-1	0	0	-1	0	0	1	1	0	0	0	0	-1	1	-1
UK	0	1	-1	0	1	-1	1	-1	-1	0	0	0	0	1	-1
Luxembourg	1	0	0	0	0	0	0	1	0	0	0	0	-1	-1	1
Spain	0	-1	0	0	-1	0	1	-1	-1	1	0	0	-1	-1	1
Germany	0	-1	0	0	-1	0	1	1	-1	0	0	-1	-1	1	0
Finland	-1	0	0	-1	0	0	1	-1	-1	0	0	0	0	0	-1
Denmark	-1	0	0	-1	0	0	1	1	-1	0	0	0	-1	0	-1
Netherlands	0	0	0	0	0	0	1	1	-1	0	0	0	-1	1	1
Ireland	-1	-1	-1	0	0	-1	1	0	0	0	0	0	-1	0	1
Poland	1	1	-1	0	1	-1	1	1	-1	0	0	-1	1	0	-1
Hungary	1	-1	-1	0	-1	-1	1	-1	-1	0	-1	-1	-1	-1	0
Romania	1	-1	0	0	0	0	1	1	0	1	0	0	-1	1	-1
Lithuania	1	1	0	1	1	0	1	-1	0	1	0	0	1	1	0
Malta	1	0	-1	1	0	0	1	1	-1	1	0	-1	na	na	na
Bulgaria	1	-1	1	1	-1	1	1	-1	-1	0	-1	-1	0	0	1
Estonia	-1	-1	0	0	0	0	1	1	0	0	0	0	-1	0	0
Croatia	na	na	0	na	na	0	1	1	0	0	0	0	0	na	na
Latvia	1	0	1	1	0	0	-1	1	0	0	1	0	na	na	na
Slovakia	0	-1	-1	0	-1	-1	-1	0	0	0	0	0	0	0	1
Slovenia	0	0	1	0	0	0	na	na	na	na	na	na	0	0	-1
Cyprus	0	-1	1	0	0	1	1	1	-1	1	1	-1	-1	0	0
Czech Rep.	1	1	0	0	1	0	0	0	-1	0	0	-1	1	1	1
improvement	14	4	5	7	4	3	21	15	0	5	2	0	3	11	10
residual	8	9	17	17	15	20	3	3	11	22	23	20	11	10	7
Decline	5	14	6	3	8	5	3	9	16	0	2	7	12	4	8

Notes:  
1/0/-1 = improvements/non-significant/declines of democracy  
left-/right-aligned country names = old/new EU member states

Source:  
Own Calculations based on Smolka (2021), Democracy Matrix V4 (Lauth und Schlenkrich 2021) and Democracy Barometer V7 (Engler u. a. 2020)

Fig. 1 Declines of Democracy—A Replication and Recalculation

the recalculation, it is noticeable that the DeMaX trends are now contoured in such a way that they only run in one direction, while the trends based on Smolka’s data still show opposing developments.

Getting back to the comparison of old and new member states, it is striking in the DeMaX analysis that only two developments in the group of old EU states are still recorded as relevant by the classification rule, indicating that minor changes may be present, but are not (yet) substantial. Compared to the findings without a threshold, the center of a potential decline of democracy clearly shifted to the new EU member states (IoD/DoD=6/8), while the old ones stand out for their stability (1/1). In contrast, the recalculation of the Smolka data again show that declines of democracy outnumber improvements among the old member states (5/10), whereas the ratio is reversed for the new member states (9/6) and improvements represent

a slight surplus compared to declines. That partially contradicts the debate, especially in recent years, which attempts to locate the epicenter of de-democratization in Eastern Europe.

In the last column of Fig. 1 a replication of the Smolka classification rule for the original Democracy Barometer dataset was performed.<sup>13</sup> Even though the absolute number of improvements and declines of democracy agree, the results based on the original data are in sharp contrast. The trends are reversed now: Whereas the analysis of the Smolka data showed a surplus of improvements in the first period followed by higher numbers of declines in the subsequent periods, the original data demonstrate that declines outnumber improvements of democracy by far in the first period, whereas the trend changes and becomes positive in the second period. This is somewhat surprising, since both analyses were carried out on the same dataset, which is why it stands to reason that Smolka's (2021) standardization method has an enormous impact on the results. Thus, these described trends could be rather unreal and an artificial product of methodological decisions.

The contrasts can be highlighted by single country comparisons: Whereas the Smolka data indicates Ireland's continuous decline of democracy across all three periods, it improves and then stagnates according to the DeMaX data, but shows a down and up for the original DB data (dynamics based on Smolka rule). To my knowledge, there is no publication that counts Ireland among the backsliders in terms of loss of procedural democratic quality. Furthermore, according to the Smolka and the original DB data, Bulgaria experienced an improvement in democracy in the period between 2012 and 2016, whereas the DeMaX shows a decline of democracy. The latter is more comprehensible since 2014 marked the return of the GERB party and its leader Borrisov to government office, a leader who is accused of being involved in corruption and of having played a major role in undermining democratic institutions (Dawson and Hanley 2016, pp. 23–25).

It is difficult to understand why the quality of democracy should have changed so regularly, and why it continually fluctuated in the form of increases followed by decreases, and why the concentration of declines of democracy in Eastern Europe contradicts the majority opinion of the debate. Therefore, I think that the face validity is lowest for the Smolka data. The fact that the democracy matrix already indicates a high degree of stability in the Western democracies when the lax threshold is set and that the focus of declines shifts to Eastern Europe seems more consistent with the literature. Moreover, the electoral rise of populists, which only peaked after the so-called refugee crisis in 2014, is cited as a central cause of the spreading dynamics of de-democratization, which is reflected in the trend of the democracy matrix, but not by the other measures. The bottom line is that the measurements differ in terms of their trends, but also in the assessments of specific countries. Thus, experts with more in-depth case knowledge must decide which measurement is more valid.

---

<sup>13</sup> By the time the manuscript was completed, the author could not find any explanation for why Smolka has fewer missing cases in her analysis than the original dataset of the democracy barometer.



### 3 How is the quality of democracy in the EU developing?

Do measurement instruments converge or diverge? The following analysis extends the study and covers several prominent democracy measures, which take a gradualist perspective and cover the relevant sample regarding the scope of countries and years.<sup>14</sup> On the one hand, we compare the trends for countries over time to identify if measurement instruments agree regarding the countries' development dynamics (diachronic). On the other hand, we compare the assessments for countries at several points in time to find out if measurement instruments assign different levels of quality of democracy to them (synchronic).

We will start with the diachronic comparison (Fig. 2). First of all, it has become obvious that *Polity* is by far more static than the other measurements and thus not appropriate to capture gradual changes within the group of democracies. It is worth noting that the few losses in the quality of democracy that are found tend to affect the old member states more: Belgium (−2) and France (−1) are affected in the first period, and the UK (−2) after 2012. For the new EU member states, with the exception of Slovakia (−1), no deterioration at all is identified. Even though they have a coarser granulation than the gradualist measurements, this only partly explains why Hungary and Poland stay in the same category over the whole period of investigation, which is a strong contradiction to common thought (Ágh 2016; Bogaards 2018).

The *Freedom in the World* (FIW) index shows opposing trends for the first period, whereby the improvements slightly outnumber the declines of democracy (6/4). Afterwards the picture changes, and a trend of decline of democracy is reported, which covers different countries in the two periods—only Hungary is devalued twice. According to FIW two thirds of the old (10/15) and about the half (7/13) of the new EU member states experience a decline at some point in time. In addition, the negative trend already starts in the old member states in the first period and persists, while five new EU members—especially Romania (+11)—made progress in the same period. Bulgaria is the only one heading the other direction. It should be added that, with the exception of Hungary, these changes are not intense. This complies with the study by Bakke and Sitter (2020, p. 11), who highlighted that Hungary and Poland are backsliding more significantly than the Czech Republic or Slovakia based on FIW data. We could summarize this by saying that a slight decline of democracy can be found in the group of old and new member states, but the low intensity of the changes also points to stability.

The *World Governance Indicators* offer the indicator of *Voice and Accountability* (WGI/VAE), which comes close to the concept of electoral democracy. The VAE displays opposing developments of quality of democracy in all three periods, whereby improvements are clearly outnumbered by declines in the first period (4/18) and are then equal in the second (9/9) and are almost equal in the third (4/7) period. It shows the second highest number of substantial changes among all measurements

<sup>14</sup> For this reason, dichotomous measurements like BNR and Regimes of the World (RoW) are not included as well as the Bertelsmann Transformations Index (BTI) or the Sustainable Governance Indicators (SGI), which either do not cover the established or young democracies or the complete period of interest.

and it also shows the Scandinavian countries experiencing ups and downs, which is a prime example of the seemingly strange fluctuation of values. Once again, the epicenter of the decline of democracy is the old EU member states.<sup>15</sup>

Since the *Electoral Democracy Index* (EDI) is the core democracy index of V-Dem and the DeMaX is a customized version of V-Dem, they show a similar trend that only varies regarding the intensity, which is why we describe them together. The measurements see an increase of democracy in the first period that weakens in the second and is followed by a decline in the third period, whereby the EDI (25/18) not only reports higher intensities of changes than the DeMaX (12/12), but also more improvements than declines in summary. Most country differences are due to the fact that one measurement tool indicates a substantial change based on the 1% threshold (often the EDI), while the other does not, but the change still points in the same direction.<sup>16</sup> Regarding the comparison of trends among old and new member states, the two measurements also agree, but the EDI shows a higher surplus of improvements for the old member states in the first period (7/0), whereas the DeMaX draws a picture of stagnation (1/0). For the new member states the ratio is turned upside down during the same period, since the surplus of improvements in the EDI (5/2) is lower than for the DeMaX (6/0). Concerning the ratio of improvements and declines of democracy in the third period, both measurements are almost identical, even though the EDI records more substantial changes. In conclusion, the improvements of democracy among the old member states from the first period are not overwhelmed from the declines in the third period according to the EDI, while the declines following the stagnation result in a slightly overall negative trend for the DeMaX. In contrast, both measurements agree that the declines did not outnumber the improvements in the first period, but the intensity of declines is higher than the improvements of the early post-accession phase among the new EU members (e.g. Malta and Poland). In addition, there are some countries with negative trends (foremost Hungary, but also the Czech Republic and Bulgaria).

Since the description of the *Democracy Barometer* (DB) data is a repetition based on the one-dimensional value of the highest level of aggregation, I decided to abbreviate this. In the first period, declines dominate (2/17), which are almost completely reversed by a similar trend in favor of improvements in the second period (16/4), which then weakens in view of an equilibrium between declines and improvements (14/10). The overall trend for the old and new member states does

<sup>15</sup> Birdwell et al. (2013, p. 98) found on the basis of the WGI indicators Political Stability, Rule of Law and Control of Corruption combined with the electoral turnout that Eastern European countries account for the six most significant improvers between 2000 and 2011, whereas the backsliders are mostly comprised of Southern- and Western European countries with the exception of Hungary. That supports the picture found here, but we also have to point out that these indicators do not measure what they are supposed to because they do not capture the very core of electoral democracy—the electoral regime and the space for political competition.

<sup>16</sup> E.g. the EDI signals substantial declines for Slovakia and Latvia in the period between 2004 and 2008, whereas the DeMaX shows declining values for these cases as well, but below the threshold of 1% of the total scale. There are a few exceptions where the direction of change between the measurement instruments is contradictory, but in none of the cases are both values classified as substantial. E.g. the EDI indicates a substantial decrease for Ireland between 2012 and 2016, whereas the DeMaX reports a non-substantial increase.

	Polity		Polity		FIW		FIW		WGI/VAE		WGI/VAE		EDI		EDI		De Max		De Max		DB		DB	
	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16	04-08	08-12	12-16
Belgium	-10	0	0	-1	0	-2	-1.64	0.34	0.4	3	0.1	0.1	0.91	0.13	-0.27	-7.13	6.42	7.3	0.01	4.66	-5.56	6.42	7.3	
France	-5	0	0	1	-1	-5	-2.16	-0.96	-1.46	1	0.4	-0.3	0.43	0.37	-0.23	0.01	4.66	-5.56	0.01	4.66	-5.56	6.42	7.3	
Spain	0	0	0	6	1	-2	-1.99	-1.73	-0.31	1.3	0.8	-4	1.49	-0.98	-1.12	-6.09	-0.19	1.85	0.43	0.37	-0.23	0.01	4.66	
Austria	0	0	0	-2	-1	-2	-1.69	1.3	-1.6	1.3	0.1	-2.5	0.9	-0.08	-1.16	-1.34	-0.24	2.96	1.49	-0.98	-1.12	-6.09	-0.19	
Ireland	0	0	0	-1	0	-1	-0.79	-1.43	-0.46	1.1	-0.3	-1.2	0.31	-0.03	0.64	-5.02	4.76	1.38	0.31	-0.03	0.64	-5.02	4.76	
Italy	0	0	0	-2	-1	-2	-2.07	-1.64	1.69	1.9	1.2	0.1	0.62	1.37	-0.11	-3.36	7.67	-3.63	0.62	1.37	-0.11	-3.36	7.67	
Denmark	0	0	0	-2	1	-1	-3.76	1.94	-1.9	0.8	0	-0.3	0.51	0.05	-0.63	-5.94	0.1	-1.38	0.51	0.05	-0.63	-5.94	0.1	
Germany	0	0	0	-3	0	-1	-2.27	0.81	-0.43	0.2	0.1	-1.6	0.27	0.13	-1.1	-0.31	6.47	-3.97	0.27	0.13	-1.1	-0.31	6.47	
Greece	0	0	0	-1	-5	1	-3.76	-3.24	-0.46	0.9	0.3	-3.1	-0.34	0.56	-0.61	-5.21	-2.47	2.04	-0.34	0.56	-0.61	-5.21	-2.47	
Finland	0	0	0	0	0	0	-4.59	1.97	-0.99	1.2	0.3	-2.1	0.33	-0.12	-0.64	-1.9	2.36	-4.22	0.33	-0.12	-0.64	-1.9	2.36	
UK	0	0	-10	0	0	-2	-3.94	0.26	-0.63	0.8	-0.2	-1.7	0.34	-0.37	-1.8	0.22	6.98	-1.43	0.34	-0.37	-1.8	0.22	6.98	
Portugal	0	0	0	0	0	0	-3.76	-2.47	1.84	1.5	-0.7	-0.1	-0.21	-0.31	-0.03	-3.12	2.95	2.44	-0.21	-0.31	-0.03	-3.12	2.95	
Luxembourg	0	0	0	0	0	-2	-1.74	1.49	-2.14	0.1	0.4	0.1	0	0.83	0.08	-6.22	-2.19	7.92	0	0.83	0.08	-6.22	-2.19	
Netherlands	0	0	0	0	0	0	-3.06	1.76	-0.94	0.7	-0.2	1.3	0.61	0.43	-0.26	-3.43	1.64	3.14	0.61	0.43	-0.26	-3.43	1.64	
Sweden	0	0	0	0	0	0	-3.13	2.43	-1.79	0.1	0.6	-0.4	0.11	-0.05	-1.66	0.3	8.76	3.17	0.11	-0.05	-1.66	0.3	8.76	
Hungary	0	0	0	-4	-12	-12	-2.61	-3.1	-4.96	0.3	-7.5	-16.9	0.63	-8.78	-7.45	-2.48	-10.6	-0.62	0.63	-8.78	-7.45	-2.48	-10.6	
Estonia	0	0	0	2	0	-1	-0.01	0.29	1.37	0.1	0.5	2.8	0.27	0.72	1.08	-2.25	2.61	1.33	0.27	0.72	1.08	-2.25	2.61	
Cyprus	0	0	0	-1	-2	1	1.91	-0.89	0.36	2	0.5	0.6	1.87	0.45	0.39	na	na	na	1.87	0.45	0.39	na	na	
Czech R.	-5	0	0	2	0	-1	1.13	-0.9	0.79	0.6	-0.8	-4.2	0.56	-0.08	-3.13	4.28	2.58	4.61	0.56	-0.08	-3.13	4.28	2.58	
Bulgaria	0	0	0	-5	-1	-1	0.11	-2.53	0.01	0.3	-4.6	-0.4	0.49	-4.93	-0.32	-7.58	0.83	10.2	0.49	-4.93	-0.32	-7.58	0.83	
Croatia	5	0	0	2	2	1	-2.71	1.04	-0.23	2.4	7.5	-8.3	1.22	6.26	-7.39	-6.02	0.8	-1.48	1.22	6.26	-7.39	-6.02	0.8	
Latvia	0	0	0	0	-3	3	1.14	-0.07	0.96	-4	3.4	2.1	-0.76	2.93	0.61	na	na	na	-0.76	2.93	0.61	na	na	
Lithuania	0	0	0	4	-1	1	-0.46	1.14	1	0.3	1.2	-1.3	2.02	-0.12	-0.08	0.51	4.84	4.2	2.02	-0.12	-0.08	0.51	4.84	
Poland	0	0	0	1	0	-4	-0.91	1.59	-3.14	1.1	-0.1	-1.6	1.13	0.28	-11.9	6.88	2.51	-8.51	1.13	0.28	-11.9	6.88		
Malta	na	na	na	-1	-1	-1	-1.76	-1.21	0.44	1.3	1.3	-4.5	1.6	0.25	-3.93	na	na	na	1.6	0.25	-3.93	na	na	
Romania	0	0	0	11	-2	3	2.09	-2.79	3.17	1.8	4.5	6.1	6.79	0.72	7.85	-3.66	6.14	-6.18	6.79	0.72	7.85	-3.66	6.14	
Slovakia	5	0	0	1	1	-3	0.19	0.54	-0.17	-1.7	1.6	-1.3	-0.57	0.66	-0.88	-6.68	-3.56	3.58	-0.57	0.66	-0.88	-6.68	-3.56	
Slovenia	0	0	0	0	0	0	-1.36	-0.24	0.16	0.6	1.8	1.5	na	na	na	-0.48	2.82	-2.12	na	na	na	-0.48	2.82	
improve	2	0	0	6	1	2	4	9	4	12	8	5	7	3	2	2	16	14	7	3	2	2	16	
residual	22	27	26	18	21	18	6	10	17	14	18	9	20	22	15	6	5	1	20	22	15	6	5	1
decline	3	0	1	4	6	8	18	9	7	2	2	14	0	2	10	17	4	10	0	2	10	17	4	10

Notes: Values are differences expressed in percent to their scale spans. Polity = 20, FIW = 100, WGI/VAE = 7; EDI = 1; DeMax = 1; DB = 2. Dark grey shaded cells with bold numbers = decline of democracy (-1), light grey with italics = improvement of democracy (1), white cells = residual category. Source: Own compilation based on the latest versions of the datasets

Fig. 2 Comparison of Measurement Instruments

not differ, but there are more declines among the old member states. These areas also saw a higher number of improvements in the subsequent periods. Thus, the DB reports change in every direction and only seldom stability (old EU: 5 out of 45, new EU: 5 out of 30 observations).<sup>17</sup>

To illustrate the contrasts between the measurements and to do some face validity, I will briefly exemplify some country comparisons. For Estonia and Latvia, which are often described as anchors of stability (Cianetti 2018), only the DB, the FIW and the EDI show backsliding patterns. In contrast to the literature (Hanley and Vachudova 2018) that counts the Czech Republic as participant in the group of backsliders (EDI and DeMaX), the VAE and DB see the quality of democracy as having improved. Whereas the V-Dem based measurements of EDI and DeMaX do not report any substantial declines for most of the old member states like France, Belgium, Italy, Denmark, Luxembourg and so on, the other measurements do see decline at some points in time (especially DB and VAE).

Are the EU member states experiencing a decline of democracy? The results of the measurement instruments do not provide a clear answer to this question. In summary, Polity is unable to detect any trend. The VAE and DB see a decline of democracy in the first period and a weakening of the negative trend (VAE) or even improvements (DB) afterwards. In contrast, the EDI and DeMaX report an improvement in the first period that weakens and changes into a decline of democracy after 2012. The FIW comes close to the latter ones, but the improvements are less intense and the declines set in earlier. Based on the proportional trends, the DB and especially the VAE localize the epicenter of decline of democracy in the old EU (Polity and FIW as well, but only slightly), whereas the DeMaX and even more the EDI shift it to the new EU. Without having the necessary case knowledge for every country, the correspondence of the V-Dem based measurements to the literature seems to be greater. One thing that all the measuring instruments have in common is that they indicate opposite developments, so that some countries lost quality of democracy, whereas others gained it.

We continue with a short synchronic comparison to see if the measurement instruments are similar in terms of their assessments of the state of democracies at the selected points in time. Since the measurements cannot be compared directly, we rely on their rankings and rather interpretative insights.<sup>18</sup> Beginning with Polity once again, in which the EU member states vary within the three top categories of the scale (8–10 or 90–100% expressed on a relative scale). Counterintuitively, Polity cannot detect significant differences between the old and new EU member states. In addition, the UK and Belgium land in the lowest category of eight points among the EU member states, whereas Hungary and Poland remain in the top category. This is obviously not a finding that reflects the majority opinion in the research.

<sup>17</sup> Bochsler and Juon (2020, 173) conclude that the min–max transformed data of the DB do “not support the notion of an overall regional deterioration in the quality of democracy” between 1990 and 2016. Instead, they see improvements (especially in Latvia and Lithuania), some deterioration, but mainly stagnation since the 2000s and country-specific developments that neither capture all components of democracy nor can be summarized in regional patterns. This interpretation only partly overlaps with my analysis.

<sup>18</sup> Alternatively, the measurements could be z-standardized or relative scales could be constructed based on their scale spans (e.g. a value of 6 on the Polity scale would be expressed as 80% on a relative scale).

The FIW differentiates more strongly with regard to the range, since the minimum Romania with 72 points has a clear distance to the maximum of 100. However, it also shows that the old EU states, with the exception of Greece and Italy, all score well above 90. In addition, France now receives a slightly lower score than some of the new member states that moved into the top-performing group (e.g. Czech Republic, Estonia and Slovenia) in 2004. In contrast, Hungary, which started with a higher score than some of the old EU members, fell to last place in the ranking. Despite the outstanding improvement it had received from FIW, Romania still ended up in third last place in 2016.

The range of the UAE is large and is marked by Romania (0.32) on one side and Denmark (1.80) on the other. Differences are also shown within the two groups. What is most striking, however, is the clear separation between old and new EU member states, which was almost perfect in 2004 and is only interrupted at the group borders by Cyprus, Malta and Estonia or Spain, Greece and Italia in 2016. It could be added that the prominent backsliders had—in contrast to the results from the FIW—a strong difference to the established democracies in 2004.

The range of the EDI is similar to that of the FIW, ranging from Romania (0.61) to Denmark (0.92). The differences within the group of established democracies are rather small. There was no clear demarcation between old and new EU states in 2004, but in contrast to the previous measurements, the demarcation is becoming more pronounced rather than less pronounced over the three periods. It is worth mentioning that Estonia ranked third in 2016. While Hungary is often cited as a prime example of the deep fall of a former showcase state, in the EDI it is Poland. Poland even made it to seventh place in 2004 but found itself in third to last place in 2016. Furthermore, the high rank—one of the top values in the sample—for Greece in 2012 is a sharp contrast to other measurements. Less surprisingly, the DeMaX resembles many features of the EDI and especially in the way that the distinction between old and new member states becomes more evident over the course of time.

Due to the range from Slovakia (3.43) to Denmark (4.63) the sample is more tied together than for the EDI, DeMaX and FIW, but there is variance within the new and old member states, which build loose groups as has been described many times before. Since France and Portugal rank much lower in some measurements while at the same time Slovenia and Czech Republic rank higher in the others, the DB does offer some country-specific contrasts.<sup>19</sup> In summary, the EDI and DeMaX show the similarities with the FIW and VAE, whereas all of them differ from the DB.

## 4 Conclusion

Do the measurements converge or diverge? The comparison between the measurements has made it clear that they contrast strongly and partly offer completely opposite assessments. But how can we deal with these findings and what strategies can help to solve or mitigate the selection problems? First of all, the choice of mea-

---

<sup>19</sup> For a better comparison of country-specific differences across measurements, the harmonized samples should be z-standardized.

surements requires careful selection with the research question being the linchpin for the reasoning context. This does not only mean constraints in terms of temporal or geographic coverage, but the underlying conceptual meanings of measurements. For example, in the case of the debate about the development of democracy, there are several lines of interpretation that cannot be answered with one measurement instrument: The most common views focus on the quality of democracy in a sense of quality of procedures and contents. Others extend the concept and look at the stability of democracy by analyzing the sociocultural embedding of democratic institutions. Another step further are assessments of the quality of government that evaluate policy outputs and the conversion process (Schmitter 2015, p. 36).

Moreover, the sample under investigation may require different measurements: Whereas thin measurements can be applied to research questions that are concerned with a large number of countries and long time-span, they are of limited value for analyses focusing on the democratic spectrum of the regime continuum. Thus, the application of enriched measurements like liberal democracy, which include the Rule of Law and accountability patterns are more suitable since they are expected to be the Achilles' heel of young democracies and account for most of the variance among democracies. Related to that, the detection of development dynamics is also strongly dependent on the overall granulation of the measurements. As a matter of course, dichotomous measurements, but also rough scalings like the categorized freedom rating of the FIW or Polity, face a problem of a "clump effect", which appears when very different cases end up in one category, and which can often lead to the masking of relevant variance.

In addition to the adequacy, the concept validity of the measurement must of course be included in the reasoning, which precedes the results of the measurements. This strand of literature took off with the seminal article by Munck and Verkuilen (2002), and has since then seen major improvements (inconclusive list Pickel et al. 2015; Skaaning 2018; McMann et al. 2021), although there are still some unresolved issues. More importantly, however, users will rely more heavily on these studies than they have in the past when selecting an appropriate measurement tool. The results of any study carried out on the basis of the Democracy Index of the Economist Intelligence Unit should raise concerns because the measurement has a quite low concept validity.

Then, there are two possibilities if we accept the fact of contrastive measurements and try to eliminate or at least reduce doubts about validity of results without choosing one instrument over another. On the one hand, as presented here, the robustness of the analysis can be assessed by substituting democracy measures to obtain meta-analytic certainty about the uniformity of the results. Waldner and Lust (2018, p. 97) pointed out that "choice of indicators has an enormous impact on empirical findings" which is why robustness checks that replicate studies on the basis of different measurements should become a mandatory exercise in the research state. In addition, not only should the choice of measurement instruments be subjected to robustness tests, but also the classification rules and threshold settings. This does not only apply to descriptive studies, but also to causal analyses that should test

models based on different measurements of the dependent variable (or independent variable).<sup>20</sup>

On the other hand, there is the possibility—as was done in the project on Unified Democracy Scores (UDS) (Pemstein et al. 2010)—to offset the results of the measurements in order to conduct the analyses on an average assessment that reflects the variance of contrasting assessments. The advantage of the first variant—the robustness checks by replicating the results—is the preservation of information on the basis of the underlying measurements. In contrast to the second variant, this is also a disadvantage since the aggregation reduces complexity. In the end, the meta-analysis by aggregation is calculative while the meta-analysis by replication is interpretative.

On a scale, changes to equal intensity from two cases are assumed to be identical, which is also justified by the assumption of an interval scale in most gradual measurement instruments.<sup>21</sup> However, both the initial and subsequent states are important for interpreting the changes to quality of democracy: Even at identical intensities of changes, they may modify the interpretation of changes of quality of democracy. Thus, their combination enriches the debate (Stanley 2019, p. 351). On the one hand, it makes a difference if a country suffers a loss of quality of democracy (difference in degree) or experiences a breakdown of democracy (difference in kind). On the other hand, it can be assumed that changes occur more frequently in the middle of the regime continuum than at the edges, which is not to negate the stability of hybrid regimes in the gray zone. The other way around, minor changes close to a threshold should not be exaggerated. Similarly, the interpretation of change can be contextualized by the diachronic progression of a case since a change in the form of a regular fluctuation is certainly less significant than a singular deviation at an otherwise constant level. In my view, the label autocratization is misleading if it conflates developments with such different outcomes including democratic recession, hybridization, autocratic consolidation and finally, democratic breakdown (see Lührmann and Lindberg 2019).

Similarly, it is often a good piece of advice to not allow cases with minor changes from getting into the group of substantial changes (set-theoretic thinking). As a matter of course that comes at the cost of setting a threshold and defending it against criticism. Concerning the problem of setting a threshold, Skaaning (2020, p. 1535) correctly points to the confidence intervals of the Varieties of Democracy Project to integrate an evaluation of the significance of country changes. Since not all measurements deliver information about uncertainty, the key is to do robustness checks once again. If the results change significantly with small shifts of the threshold, there is

---

<sup>20</sup> For example, Petersheim's (2012, p. 84) analysis of potential factors influencing the quality of democracy was not robust to alter the dependent variable by different measurement instruments. Møller and Skaaning (2014; ch. 9) also demonstrated how the significance levels and signs of the regression coefficients vary with divergent measures of the rule of law.

<sup>21</sup> Partial criticism is also voiced about how gradual measurements cannot hold the preconditions for such a constancy assumption (Maerz et al. 2021, p. 5). To counter such measurement inaccuracies, the only way out is to systematically include the uncertainty in the form of confidence or credibility intervals. However, only a few measurement instruments offer these (V-Dem, UDS, and WGI).



clearly a high level of sensitivity and, thus, the threshold should be reconsidered or at least the issue should be addressed.

The case selection in terms of defining periods within the units of the analysis (2004–2008–2012–2016) must also be criticized, which is also a self-criticism, because I have adopted the procedure for the replication: While the starting point of 2004 as the year of the EU's Eastern enlargement can still be reasonably justified, it is questionable whether the four-year cycle is sensibly chosen, since it does not capture temporary fluctuations in between.<sup>22</sup> Therefore, developments could be kept artificially small or inflated, since the constructed periods cut up episodes of decline or improvement of democracy. A more sensible approach is the one proposed by Lührmann and Lindberg (2019), who try to identify episodes. This quasi-qualitative approach to sequencing, which better captures country-specific dynamics without compromising comparability, has also been improved in the Episodes of Regimes Transition (ERT) framework and marks an important contribution to future research efforts.

If we take the multidimensionality of the concept of democracy seriously, concepts to capture declines of democracy should not only operate at the highest level of aggregation but should take into account how many characteristics and to what extent they are affected by a decline in quality. I suggest terming the number of changed features as “extensity” and the degree of change as “intensity”. In this respect, Smolka's approach is commendable since a decline of democracy is identified as a multidimensional development. Moreover, to a certain extent, there is an offsetting of extensity and intensity in that the overall value and not just the partial indices must decline. The threshold values should be set on a dimension-oriented basis: On the one hand, the scales of the dimensions are rarely uniform neither in terms of their construction nor their underlying distribution of observations. On the other hand, the overall value can express the dimensions of a concept differently regarding varying aggregation rules, which, therefore, demands varied threshold setting for changes on overall and dimensional scales.

Linked to that is one future research challenge, the disaggregation of declines (and also improvements) of democracy in order to detect where de-democratization takes place (Gora and de Wilde 2020), where it starts<sup>23</sup> and to learn how deficits spill-over in order to unravel its endogenous dynamics. Since there are many ways to be a deficient democracy (Lauth et al. 2021), there are also many ways of how de-democratization may unfold, which is why we should go on a search for patterns that constitute dominant pathways. This was not part of the analysis of this paper, but there are already some studies in the state of research that try to differentiate the empirical analysis of improvements and declines of democracy in this way. Once

<sup>22</sup> Rupnik's (2007, p. 19) reference to the limits of linearity must be transferred from the transformation phase model to the recording of gradual developments of transformation pathways, which rarely occur in a straight line but rather in a zigzag course. In this respect, it can be considered quite problematic that the V-Dem model already tends to construct episodes and thereby potential intra-case variance is suppressed. However, this is only an observation due to face-validity and needs to be clarified by experts with greater knowledge of V-Dem's Bayesian Item-Response Theory (IRT) estimation strategy.

<sup>23</sup> With regard to declines of democracy, the rule of law and accountability structures seem to be the Achilles' heel of the Eastern European member states (Bochsler und Juon 2020, 173).



again, the disaggregation of the concept of democracy should be applied in causal analysis as well, since it is unlikely that potential factors explaining declines or improvements of democracy “have uniform effects across different dimensions of democracy” (Knutsen et al. 2019, p. 2).

It turns out that quantifying measurement instruments are generally unable to capture subtle nuances of an evolutionary manner as well as short-term and temporal changes, suggesting a linearity assumption or at least periodization in the process of data generation. It is likely that this issue will become even more important when leaving the highest levels of aggregation and moving to attributes that take the multidimensionality of democracy seriously, since quantifying measurements tend to relate the developments of different dimensions so that they run in identical directions. However, there is the possibility of asymmetric and opposing developmental dynamics, which have a low chance of being detected by quantifying measurements for this reason. For example, the Rule of Law could be improved while the electoral regime and political competition experience restrictions. In contrast, ruptures and long-term developments with substantial changes can be satisfactorily represented by quantitative measurement instruments. Based on these assumptions, this process would yield the application of qualitative methods and case knowledge when the study design is focused on nuanced as well as short-term and temporary fluctuations, whereas quantitative methods are more appropriate for the analysis of ruptures and long-term variance.<sup>24</sup>

However, the self-limitation of democracy measurements is not new, and it has often been pointed out that macro-structural analyses can only provide a starting point for further research, which should then be deepened micro-analytically. Although, the data infrastructure has improved significantly—especially since the start of V-Dem—it cannot replace qualitative research. The analysis above shows how difficult it is to assess the severity of changes, which can only be done in the end with in-depth case knowledge, which is ultimately the basis for evaluating the results of quantitative measurements. An interesting push to better evaluate the changes comes from Levitz and Pop-Eleches (2010, p. 464), who conclude that the evidence for a post-accession thesis—QoD stagnates or even shrinks once the hurdle for EU accession is met—is rather weak when the developments of the new member states are compared with their “post-communist peers”. It is likely that such a transfer, in the form of establishing relations between cases and samples or samples and the universe of cases, facilitates the evaluation of development dynamics in terms of their intensity. Moreover, it is highly likely to rule out the possibility that the change is artificial and brought about by conceptual decisions of a measurement instrument.<sup>25</sup>

Finally, the mixed results of the data analysis should not deny or downplay a loss of the quality of democracy, but I do not consider the extent and intensity of these

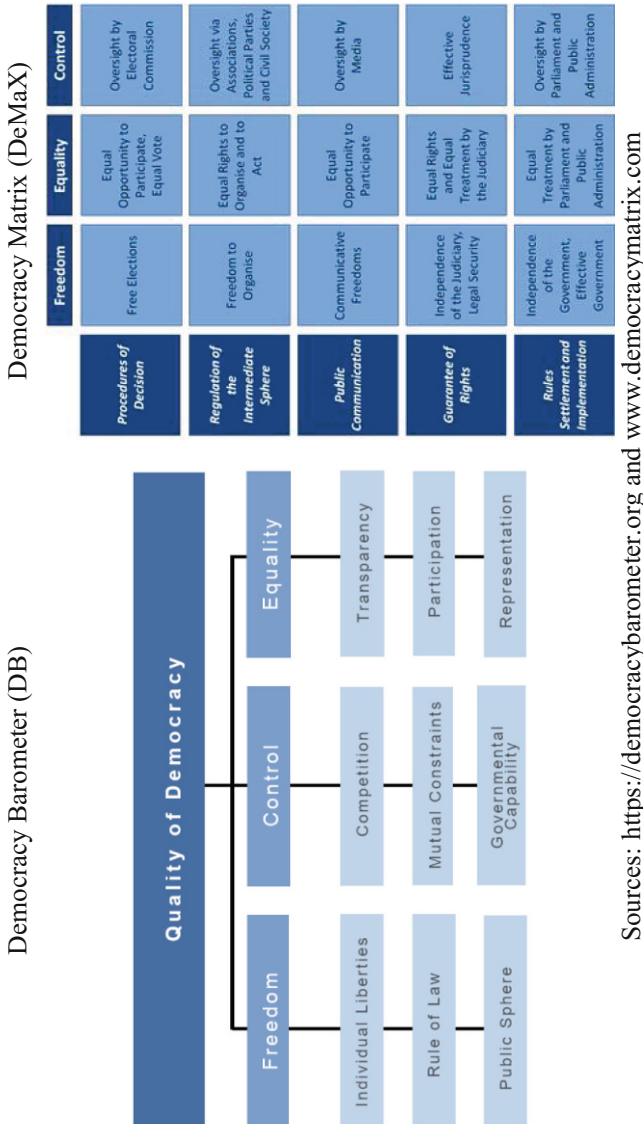
---

<sup>24</sup> Depending on the theoretical concept of the analysis, this also requires the combination of different measurement instruments to measure all components of the concept. This creates another methodological problem, as the data generated by different measurement instruments have to be standardized for further processing, with z-standardization probably offering the best solution possibility.

<sup>25</sup> The possibility that all countries are declining is rather low which is why such a pattern is more likely caused by choices of standardization methods or similar statistical procedures.

developments to be as serious as some contributions to the debate make them out to be. Especially beyond the procedural quality of liberal democracy, when we include the stability of democratic institutions in the sense of their acceptance by elites and masses as well as their performance with regard to system functions like integration or distribution, the quality of deliberation or factual participation, changes may be observed (see e.g. Kriesi 2020). Most of the countries experiencing a decline of democracy are losing ground, and other democracies are facing serious attacks, but they are still democracies, as a vast majority of the measuring instruments note. In this respect, exaggerations that declare the emergence of autocracies or propagate them as inevitable consequences are dangerous as they can undermine the fight for democracies and constitute self-fulfilling prophecies.

Appendix



**Fig. 3** Concept Trees in Comparison. **a** Democracy Barometer (DB), **b** Democracy Matrix (DeMaX). (Sources: <https://democracybarometer.org> and [www.democracymatrix.com](http://www.democracymatrix.com))



**Table 1** (Continued)

	2004–2008			2008–2012			2012–2016			Smolka rule			1% rule				
	O	F	E	C	O	F	E	C	O	F	E	C	O	F	E	C	
Slovakia	0.3	-1.0	2.0	0.0	-5.2	-4.4	-1.7	-9.5	-1.6	-10.6	2.3	-5.9	0	-1	-1	0	-1
Slovenia	3.3	-0.2	-0.6	10.6	-0.2	0.2	0.6	-1.3	2.0	-2.7	0.2	8.1	0	0	1	0	0
Spain	-0.1	-2.2	-0.5	2.6	-6.1	-6.8	-0.4	-11.1	-0.4	3.6	-5.8	1.2	0	-1	0	0	-1
Sweden	2.5	0.5	-1.6	8.4	-0.7	-7.0	2.1	2.7	4.3	0.6	8.6	3.6	1	0	1	0	0
UK	-0.1	-0.9	-2.8	3.3	3.0	3.3	2.4	3.3	-7.5	-3.8	1.9	-20.2	0	1	-1	0	1

*O* Overall Quality of Democracy, *F* Freedom, *E* Equality, *C* Control

**Table 2** Raw Data and Classifications for Democracy Matrix

	2004–2008						2008–2012						2012–2016						Smolka rule				1% Skala			
	O	F	E	C	O	F	O	F	E	C	O	F	O	F	E	C	04–08	08–12	12–16	16–16	04–08	08–12	12–16	16–16		
	Austria	-0.01	-0.02	-0.01	-0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	-1.00	-1.00	1.00	1.00	0.00	0.00	0.00	0.00	
Belgium	0.00	0.00	-0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.01	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Bulgaria	0.00	0.04	-0.04	-0.02	-0.05	-0.07	-0.04	-0.03	-0.03	-0.04	0.01	0.00	0.01	0.00	0.01	0.01	-1.00	-1.00	1.00	1.00	0.00	-1.00	0.00	-1.00		
Croatia	-0.07	-0.10	-0.07	-0.06	0.06	0.09	0.08	0.03	0.03	0.08	0.01	0.01	0.01	0.01	0.01	0.02	-1.00	1.00	1.00	1.00	-1.00	1.00	1.00	0.00		
Cyprus	0.00	0.01	0.01	-0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.02	0.03	0.03	0.00	0.03	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Czech R	-0.03	-0.02	-0.03	-0.05	0.00	0.00	0.00	-0.01	-0.01	0.00	0.01	0.01	0.01	0.01	0.01	-0.01	-1.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00		
Denmark	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	-1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00		
Estonia	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Finland	-0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	-1.00	1.00	1.00	0.00	0.00	0.00	0.00		
France	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	-1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Germany	-0.01	-0.02	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	0.00	1.00	1.00	0.00	-1.00	0.00	0.00		
Greece	-0.01	-0.01	-0.01	0.00	0.01	-0.02	0.01	-0.02	0.01	0.02	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00		
Hungary	-0.08	-0.08	-0.08	-0.07	-0.09	-0.09	-0.05	-0.05	-0.05	-0.13	0.01	0.00	0.01	0.00	0.00	0.02	-1.00	-1.00	1.00	1.00	-1.00	-1.00	1.00	-1.00		
Ireland	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00		
Italy	0.00	0.00	-0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.03	0.01	0.00	0.01	0.00	0.01	0.01	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Latvia	0.01	0.01	0.01	0.00	0.03	0.05	0.02	0.02	0.02	0.02	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	1.00	1.00	-1.00	0.00	0.00	0.00	0.00	0.00		
Lithuania	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.03	0.02	0.03	0.02	0.01	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00		
Luxembourg	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00		
Malta	-0.04	-0.04	-0.04	-0.04	0.00	0.01	0.00	0.00	0.01	0.01	0.02	0.03	0.03	0.03	0.01	0.01	-1.00	1.00	1.00	1.00	-1.00	0.00	0.00	1.00		
Netherlands	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Poland	-0.12	-0.08	-0.09	-0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	-1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00		
Portugal	0.00	0.01	0.00	-0.01	0.00	0.00	0.00	-0.01	-0.01	0.00	0.00	0.01	0.00	0.01	-0.02	0.00	0.00	-1.00	-1.00	-1.00	0.00	0.00	0.00	0.00		
Romania	0.08	0.10	0.06	0.07	0.01	0.01	0.01	0.01	0.01	0.00	0.07	0.08	0.04	0.08	0.04	0.08	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00		

**Table 2** (Continued)

	2004–2008			2008–2012			2012–2016			Smolka rule			1% Skala				
	O	F	E	C	O	F	E	C	O	F	E	C	O	F	E	C	
Slovakia	-0.01	0.00	-0.03	0.01	0.01	0.01	0.02	-0.01	-0.01	-0.01	-0.01	0.00	0.00	-1.00	0.00	0.00	0.00
Slovenia	NA	0.01	0.04	NA	NA	0.01	0.01	NA	NA	0.01	-0.01	NA	NA	NA	NA	NA	NA
Spain	-0.01	-0.01	-0.02	-0.01	-0.01	0.00	-0.02	-0.02	0.02	0.02	0.01	0.01	-1.00	-1.00	0.00	0.00	1.00
Sweden	-0.02	-0.02	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00	0.00
UK	-0.02	-0.02	-0.03	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.01	0.00	-1.00	-1.00	0.00	0.00	0.00

O Overall Quality of Democracy, F Freedom, E Equality, C Control

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ágh, Attila. 2016. The decline of democracy in east-central Europe. *Problems of Post-Communism* 63:277–287. <https://doi.org/10.1080/10758216.2015.1113383>.
- Ariely, Gal, and Eldad Davidov. 2012. Assessment of measurement equivalence with cross-national and longitudinal surveys in political science. *European Political Science* 11:363–377. <https://doi.org/10.1057/eps.2011.11>.
- Bakke, Elisabeth, and Nick Sitter. 2020. The EU's enfants terribles: democratic backsliding in central Europe since 2010. *Perspectives on Politics* <https://doi.org/10.1017/S1537592720001292>.
- Birdwell, Jonathan, Sebastian Feve, Chris Tryhorn, and Natalia Vibla. 2013. *Backsliders: measuring democracy in the EU*. London: Demos.
- Bochsler, Daniel, and Andreas Juon. 2020. Authoritarian footprints in Central and Eastern Europe. *East European Politics* 36:167–187. <https://doi.org/10.1080/21599165.2019.1698420>.
- Bogaards, Matthijs. 2018. De-democratization in Hungary: diffusely defective democracy. *Democratization* 25:1481–1499. <https://doi.org/10.1080/13510347.2018.1485015>.
- Cianetti, Licia. 2018. Consolidated technocratic and ethnic hollowness, but no backsliding: reassessing Europeanisation in Estonia and Latvia. *East European Politics* 34:317–336. <https://doi.org/10.1080/21599165.2018.1482212>.
- Cianetti, Licia, and Seán Hanley. 2021. The end of the backsliding paradigm. *Journal of Democracy* 32:66–80. <https://doi.org/10.1353/jod.2021.0001>.
- Cianetti, Licia, James Dawson, and Seán Hanley. 2018. Rethinking “democratic backsliding” in Central and Eastern Europe—looking beyond Hungary and Poland. *East European Politics* 34:243–256. <https://doi.org/10.1080/21599165.2018.1491401>.
- Dawson, James, and Seán Hanley. 2016. The fading mirage of the “liberal consensus”. *Journal of Democracy* 27:20–34. <https://doi.org/10.1353/jod.2016.0015>.
- Diamond, Larry Jay, and Leonardo Morlino. 2004. Quality of democracy: an overview. *Journal of Democracy* 15:20–31. <https://doi.org/10.1353/jod.2004.0060>.
- Doorenspleet, Renske. 2000. Reassessing the three waves of democratization. *World Politics* 52:384–406.
- Engler, Sarah, Lucas Leemann, Tarik Abou-Chadi, Heiko Geibler, Karima Bousbah, Daniel Bochsler, Marc Bühlmann, et al. 2020. *Democracy barometer*. Version, Vol. 7. Aarau: Zentrum der Demokratie.
- Erdmann, Gero. 2011. Decline of democracy: loss of quality, Hybridisation and breakdown of democracy. In *Regression of democracy?*, ed. Gero Erdmann, Marianne Kneuer, 21–58. Wiesbaden: VS. [https://doi.org/10.1007/978-3-531-93302-3\\_2](https://doi.org/10.1007/978-3-531-93302-3_2).
- Fuchs, Dieter, and Edeltraud Roller. 2018. Conceptualizing and measuring the quality of democracy: the citizens' perspective. *Politics and Governance* 6:22–32. <https://doi.org/10.17645/pag.v6i1.1188>.
- Gora, Anna, and Pieter de Wilde. 2020. The essence of democratic backsliding in the European Union: deliberation and rule of law. *Journal of European Public Policy* <https://doi.org/10.1080/13501763.2020.1855465>.
- Hanley, Seán, and Milada Anna Vachudova. 2018. Understanding the illiberal turn: democratic backsliding in the Czech Republic. *East European Politics* 34:276–296. <https://doi.org/10.1080/21599165.2018.1493457>.
- Huntington, Samuel P. 1991. Democracy's third wave. *Journal of Democracy* 2:12–34. <https://doi.org/10.1353/jod.1991.0016>.



- Jäckle, Sebastian, Uwe Wagschal, and Rafael Bauschke. 2012. Das Demokratiebarometer: “basically theory driven”? *Zeitschrift für Vergleichende Politikwissenschaft* 6:99–125. <https://doi.org/10.1007/s12286-012-0133-6>.
- Jäckle, Sebastian, Uwe Wagschal, and Rafael Bauschke. 2013. Allein die Masse macht’s nicht – Antwort auf die Replik von Merkel et al. zu unserer Kritik am Demokratiebarometer. *Zeitschrift für Vergleichende Politikwissenschaft* 7:143–153. <https://doi.org/10.1007/s12286-013-0148-7>.
- Knutsen, Carl Henrik, and Simone Wegmann. 2016. Is democracy about redistribution? *Democratization* 23:164–192. <https://doi.org/10.1080/13510347.2015.1094460>.
- Knutsen, Carl Henrik, John Gerring, Svend-Erik Skaaning, Jan Teorell, Matthew Maguire, Michael Coppedge, and Staffan I. Lindberg. 2019. Economic development and democracy: an electoral connection. *European Journal of Political Research* 58:292–314. <https://doi.org/10.1111/1475-6765.12282>.
- Kriesi, Hanspeter. 2020. Is there a crisis of democracy in Europe? *Politische Vierteljahresschrift* 61:237–260. <https://doi.org/10.1007/s11615-020-00231-9>.
- Lauth, Hans-Joachim. 2010. Möglichkeiten und Grenzen der Demokratiemessung. *Zeitschrift für Staats- und Europawissenschaften* 8:498–529.
- Lauth, Hans-Joachim, and Oliver Schlenkrich. 2018. Making trade-offs visible: theoretical and methodological considerations about the relationship between dimensions and institutions of democracy and empirical findings. *Politics and Governance* 6:78–91. <https://doi.org/10.17645/pag.v6i1.1200>.
- Lauth, Hans-Joachim, and Oliver Schlenkrich. 2021. *Democracy matrix. Version V4*. Würzburg: Democracy Matrix (DeMaX).
- Lauth, Hans-Joachim, Oliver Schlenkrich, and Lukas Lemm. 2021. Different types of deficient democracies: reassessing the relevance of diminished subtypes. *International Political Science Review* <https://doi.org/10.1177/0192512121995686>.
- Levitsky, Steven, and Lucan Way. 2015. The myth of democratic recession. *Journal of Democracy* 26:45–58. <https://doi.org/10.1353/jod.2015.0007>.
- Levitz, Philip, and Grigore Pop-Eleches. 2010. Why no backsliding? The European Union’s impact on democracy and governance before and after accession. *Comparative Political Studies* 43:457–485. <https://doi.org/10.1177/0010414009355266>.
- Lührmann, Anna, and Staffan I. Lindberg. 2019. A third wave of autocratization is here: what is new about it? *Democratization* 26:1095–1113. <https://doi.org/10.1080/13510347.2019.1582029>.
- Maerz, Seraphine F., Amanda Edgell, Matthew C. Wilson, Sebastian Hellmeier, and Staffan I. Lindberg. 2021. *A framework for understanding regime transformation: introducing the ERT Dataset*. V-Dem Working Paper Series, Vol. 113. Gothenburg: University of Gothenburg. <https://doi.org/10.2139/ssrn.3781485>.
- McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell, and Staffan Lindberg. 2021. Assessing data quality: an approach and an application. *Political Analysis* <https://doi.org/10.1017/pan.2021.27>.
- Merkel, Wolfgang, Dag Tanneberg, and Marc Bühlmann. 2013. “Den Daumen senken”: Hochmut und Kritik: Eine Replik auf die Kritik des Demokratiebarometers von Jäckle, Wagschal und Bauschke. *Zeitschrift für Vergleichende Politikwissenschaft* 7:75–84. <https://doi.org/10.1007/s12286-013-0145-x>.
- Møller, Jørgen, and Svend-Erik Skaaning. 2013. The third wave: inside the numbers. *Journal of Democracy* 24:97–109. <https://doi.org/10.1353/jod.2013.0057>.
- Møller, Jørgen, and Svend-Erik Skaaning. 2014. *The rule of law*. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137320612>.
- Mounk, Yascha. 2020. The end of history revisited. *Journal of Democracy* 31:22–35. <https://doi.org/10.1353/jod.2020.0002>.
- Munck, Gerardo L. 2016. What is democracy? A reconceptualization of the quality of democracy. *Democratization* 23:1–26. <https://doi.org/10.1080/13510347.2014.918104>.
- Munck, Gerardo L., and Jay Verkuilen. 2002. Conceptualizing and measuring democracy: evaluating alternative indices. *Comparative Political Studies* 35:5–34. <https://doi.org/10.1177/001041400203500101>.
- Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. Democratic compromise: a latent variable analysis of ten measures of regime type. *Political Analysis* 18:426–449. <https://doi.org/10.1093/pan/mpq020>.
- Petersheim, Meredith-Joy. 2012. The European Union and consolidating democracy in central and eastern Europe. *Journal of European Integration* 34:75–91. <https://doi.org/10.1080/07036337.2011.638064>.
- Pickel, Susanne, Toralf Stark, and Wiebke Breustedt. 2015. Assessing the quality of quality measures of democracy: a theoretical framework and its empirical application. *European Political Science* 14:496–520. <https://doi.org/10.1057/eps.2015.61>.

- Pickel, Susanne, Wiebke Breustedt, and Theresia Smolka. 2016. Measuring the quality of democracy: why include the citizens' perspective? *International Political Science Review* 37:645–655. <https://doi.org/10.1177/0192512116641179>.
- Pinheiro, Paulo Sérgio. 1999. The rule of law and the underprivileged in Latin America: introduction. In *The (un)rule of law and the underprivileged in Latin America*, ed. Juan E. Méndez, Guillermo O'Donnell, and Paulo Sérgio Pinheiro, 1–18. Notre Dame: University of Notre Dame Press.
- Puddington, Arch. 2008. Is the tide turning? *Journal of Democracy* 19:61–73. <https://doi.org/10.1353/jod.2008.0021>.
- Rupnik, Jacques. 2007. Is east-central Europe backsliding? From democracy fatigue to populist backlash. *Journal of Democracy* 18:17–25.
- Schmitter, Philippe C. 2015. Crisis and transition, but not decline. *Journal of Democracy* 26:32–44. <https://doi.org/10.1353/jod.2015.0004>.
- Skaaning, Svend-Erik. 2018. Different types of data and the validity of democracy measures. *Politics and Governance* 6:12.
- Skaaning, Svend-Erik. 2020. Waves of autocratization and democratization: a critical note on conceptualization and measurement. *Democratization* 27:1533–1542. <https://doi.org/10.1080/13510347.2020.1799194>.
- Smolka, Theresia. 2021. Decline of democracy—the European Union at a crossroad. *Zeitschrift für Vergleichende Politikwissenschaft* 15:81–105. <https://doi.org/10.1007/s12286-021-00481-w>.
- Stanley, Ben. 2019. Backsliding away? The quality of democracy in central and eastern Europe. *Journal of Contemporary European Research* 15:343–353. <https://doi.org/10.30950/jcer.v15i4.1122>.
- Waldner, David, and Ellen Lust. 2018. Unwelcome change: coming to terms with democratic backsliding. *Annual Review of Political Science* 21:93–113. <https://doi.org/10.1146/annurev-polisci-050517-114628>.