BIAS IN THE EYE OF BEHOLDER? 25 YEARS OF ELECTION MONITORING IN EUROPE

MICHAL MOCHTAK[1], ADAM DRNOVSKY[2], CHRISTOPHE LESSCHAEVE[3]

[1] Institute of Political Science, University of Luxembourg, Maison des Sciences Humaines, 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg. ORCID: 0000-0001-5598-5642; E-mail: michal.mochtak@uni.lu (*corresponding author*).
[2] Institute of Public Law and International Law, Bundeswehr University Munich, Werner-Heisenberg-Weg 39, 85579 Neubiberg, Germany. E-mail: adam.drnovsky@unibw.de.
[3] Institute of Political Science, University of Luxembourg, Maison des Sciences Humaines, 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg. ORCID: 0000-0002-0632-6760; E-mail: christophe.lesschaeve@uni.lu; E-mail: christophe.lesschaeve@uni.lu.

ABSTRACT

Building on the original corpus of OSCE monitoring reports, the paper analyzes quarter of century of election monitoring in Europe and assesses the congruence of OSCE written assessments with expert views. We show that, overall, the OSCE monitoring reports are highly correlated and congruent with expert assessments. More importantly, the level of congruence between the two increases with time. However, we also identify various forms of biases rooted in strategic interests and institutional preconditions. Mainly, we show that OSCE has a strong and positive bias towards Russia and its allies when it comes to election assessments indicating defensive and lenient stances. We theorize this mechanism as a *pushback effect* and show that although Russia's effort to cripple the activities of OSCE in the past two decades was not successful, OSCE was effectively forced into a defensive position producing less critical assessments than reality warrants.

NOTE
The article is also available open access at:
https://www.tandfonline.com/doi/full/10.1080/13510347.2021.2019219

BIAS IN THE EYE OF BEHOLDER? 25 YEARS OF ELECTION MONITORING IN EUROPE

Introduction

In the past three decades, election monitoring has become such an important factor in regimes' credibility that even authoritarian elites have started to feel obliged to invite international observers mimicking the effort of fulfilling their democratic commitments.[1] In an environment where media, governments, and international organizations listen carefully to what election monitors have to say, the official monitoring reports have increasingly affected countries' international outlooks, leading to various political as well as economic ramifications. With this much influence, international monitors have started to be dragged into thorny political entanglements often accompanied by accusations of political bias questioning the overall integrity of the monitoring missions and their goals.[2] How is this reflected in the monitoring practices in Europe? What kind of bias (if any) does prevail in a region with comparatively rich history of election monitoring, varying democratic qualities, and (sometimes) unjustified superiority complex?

The goal of the article is to assess 25 years of election monitoring in Europe as conducted by the Organization for Security and Cooperation in Europe (OSCE) and its Office for Democratic Institutions and Human Rights (ODIHR). The aim is to explore whether reports produced by OSCE contain any sort of bias and whether this bias is systematically present in the evaluation of elections in certain contexts. As such, the paper intellectually builds on the seminal work of Judith Kelley[3], yet goes beyond the original period covered while analyzing full-fledged final reports and introducing new methods in the study of election monitoring. To this end, we use the *wordscores* scaling algorithm with guided bootstrap sampling in order to

analyze positions of 303 monitoring reports, counting over 8700 pages of raw text, on a latent scale of *free and fair election* and explore how they fare against the internationally accepted standards. Moreover, the focus on OSCE/ODIHR explores the relevance of election bias in a context where election monitoring has a long tradition and has gone through a well-documented development. As such, the history of OSCE/ODIHR election monitoring tells the story of election monitoring in post- Cold War Europe and the dynamics that accompany it.

When it comes to existing literature, scholars have identified number of factors potentially driving the biased assessments of international monitors ranging from political, to economic, and strategic motivations.[4] In almost all of these settings, the observing authority is presented as the one with the upper hand following its political, economic, or strategic goals. While evaluating relevance of these assumptions in European context, the paper explores an existence of a specific type of reverse mechanism, under which international observers (OSCE) with high credibility are systematically pushed by the party being monitored to a more submissive position producing potentially favorable reports. The paper conceptualizes this mechanism as a *pushback effect* and shows how it works in the context of Russian pressure on OSCE/ODIHR monitoring activities.

The overall results show that the OSCE monitoring reports are highly correlated and congruent with independent expert views, which validates the methodologies employed and their common reference to the universally recognized standards of free and fair elections. More importantly, the level of congruence between the two increases with time. However, we also identify various forms of biases rooted in strategic interests and institutional preconditions. We find that higher GDP, GDP proportion of total natural resources rents, Official development assistance (ODA), and legislative elections are associated with more positive assessments. On the other hand, the size of the observation mission is associated with a more negative

assessment. More importantly, we show that the OSCE has a strong and positive bias towards Russia and its allies indicating persistent defensive and lenient stances. We theorize that although Russia's effort to cripple the election monitoring activities of OSCE/ODIHR in the past two decades was not successful, OSCE was effectively *pushed* into a defensive position producing less critical assessments towards some of the post-soviet countries than reality warrants. This *pushback effect* presents an additional perspective on election monitoring, its biases, and the underlying drivers explaining them. As such, our paper contributes to the literature on election monitoring, election observer bias as well as power relations in European context.

International Election Monitoring and its Contested Bias

As election monitoring has started to play a prominent role in international acceptance of all sorts of regimes, scholars as well as practitioners raised important questions concerning their impact and credibility.[5] Often, tensions have been highlighted between the proclaimed aim to improve elections through reliable and accurate assessments and the realities of balancing this goal with other objectives.[6] Specifically, the concern has been voiced that election assessments are more positive or negative than reality merits in order to serve certain (geo)political, security, and economic goals.[7] Kavakli and Kuhn[8] even argue that the calculus of outside observers depends not only on who they wish to see in power, but also who they want to keep from power.

Scholars and practitioners of election monitoring agree that independence and impartiality are the hallmarks of a good election monitoring body affecting both its credibility and positive influence. The authority enjoyed by such bodies rests on their adherence to the highest

standards of accurate and unbiased election monitoring.[9] Nevertheless, election monitors often face numerous practical obstacles that may hinder their ability to report on elections accurately. For instance, due to often-limited resources, decisions have to be made on how many observers can be deployed, how many interlocutors they can speak to, how many polling stations they can visit, where to visit them, and for how long they can stay in the country. The existing literature suggests that this has led to a disproportional monitoring of urban polling stations at the expense of stations situated in remote and rural areas.[10] Additionally, large countries often host proportionally fewer observers than smaller ones, poisoning the statistical significance of the sample of visited polling stations. Some authors also argue that cultural factors such as the observers' nationalities may have an influence on assessments made in the field.[11] Relatively overlooked remains the effect of the hosting country counter-actions which might range from diplomatic squabbles to strategic threats. In the context of OSCE monitoring missions, this "pushback" behavior is most often associated with post-soviet countries lead by Russian Federation, which has been criticizing OSCE monitoring missions since late 1990s.[12]

That said, it must be noted that significant efforts have been made to overcome these shortcomings in reaction to the rising competition among credible international monitoring actors who have started to find themselves under an increasing risk of harming their reputation and effectiveness by inaccurate assessments. In short, monitoring organizations that lack credibility also lack influence.[13] As a result, there has been a substantial increase in adherence to universally accepted principles for international election observation and codes of conduct for election observers.[14] Moreover, more sophisticated observation methodologies have been developed, aimed at improving the reliability of election assessments in general, often with a contribution and feedback from independent electoral experts.[15] As a result, well-established actors with transparent observation methodology, such as EU, OSCE, or The Carter Center

are believed to produce election observation reports that are increasingly accurate and objective.[16] The proposed mechanism has been further accelerated with an increase in the number of international election monitoring bodies and greater emphasis on the importance of credibility of the international standards.[17] Building on these theoretical claims, we assume that the scholars' assessments regarding the observed improvements should be mirrored in textual data as well, providing evidence that international standards are indeed the basis of the written monitoring reports and, more importantly, that the adherence of the monitoring reports to these international standards increases over time.[18] This leads us to our first hypothesis:

H1: The congruence between OSCE election monitoring reports and experts' views strengthens over time.

While improvements to the methodology can reduce internal sources of bias originating from, for example, a lack of resources or the nationality of the monitors, they are less effective in shielding monitoring bodies from political pressure. These pressures can come from host governments, third countries, or member states in the case of inter-governmental election monitoring organizations (IGOs) such as the OSCE. Kelley[19] points out that while most of the time election monitors provide genuine and uncontested assessments, the political and economic relationships between the monitored country and the member/funding countries of the monitoring organizations may influence the assessments. This is particularly the case of countries that are recipients of aid or military/trade partners of sponsoring states.[20]

In our case, arguably, the OSCE represents an IGO mostly dominated by the West, if not on the whole, then at least in the human rights and democracy promotion activities of the organization, including election monitoring. This claim leans on the fact that, to a large extent, the democracy promotion activities are politically, financially, and personally supported by

5

countries integrated or closely associated to the Western structures such as European Union (EU) or NATO, making up a majority among OSCE states.[21] A notable part of this dominance, besides possible political leverage of these governments within the organization, lies in the overwhelming number of staff working in the OSCE/ODIHR election monitoring missions originated from EU and NATO member states or other closely associated countries.[22] Arguably, this establishes a link through which some OSCE member states may impose leverage over activities of the OSCE in the area of election monitoring or at least create incentives and channels for socialization to certain norms. This inherent political bias then may affect favoring a set of (geo)political interests within the organization which does not have to be accepted by all member states.[23] This may range from geopolitical interests in countries such as Georgia or Ukraine and their role in regional security systems or economic interests in countries such as Azerbaijan with its vast natural resources.

We merge these theoretical expectations with real world dynamics of OSCE monitoring which over the years has been challenged multiple times, yet only one line of criticism has prevailed almost throughout the whole period under study – the allegations of political bias against Russian Federation and its allies. When it comes to Russia, the post-Cold War era has been increasingly affected by the West/East divide fueling confrontation in political, economic, as well as military arenas.[24] It has become a standard procedure that monitoring of elections in countries with strong ties to Russia or Russia itself are contested on political grounds accompanied by allegations of unfair treatment.

The situation got worse with diplomatic feud that dates back to the aftermath of the color revolutions in the 2000s when the OSCE/ODIHR played an important role in uncovering election frauds in some of the monitored states, thus contributing to the public mobilization against the non-democratic regimes.[25] The divergent opinions on election monitoring have

been voiced by Russia, however, at least since 1999 criticizing the OSCE for privileging the human rights dimension over other principles.[26] Specific objections to the OSCE election observation started to be raised in 2003 with a document prepared by delegations of Russia, Belarus, Kazakhstan, and Kyrgyzstan. It was a reaction to an "apparent intrusion" of OSCE practices and institutions, including election observation, into the internal affairs of the participating states.[27] With color revolutions and the consequences they had in the post-soviet region, a coalition of post-soviet countries led by Russia Federation started to contest how OSCE/ODIHR operated systematically. Russian rhetoric intensified and demands turned to an overhaul of OSCE election observation and its basic principles. In 2004, it led to a common declaration by the presidents of Armenia, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Ukraine, and Uzbekistan (later endorsed also by Turkmenistan) condemning the OSCE/ODIHR election observation practices and accusing OSCE of applying "double standards".[28]

The initiative, later transforming to another open declaration known as "Astana Appeal" and its successors, represents the most systematic attempt to question the integrity of election monitoring in Europe. Although other states may have questioned the OSCE/ODIHR election assessments occasionally, none of them has transformed the criticism to actual coordinated policy. Although the effort to modify the core focus of monitoring missions eventually failed and number of states reinterpreted their support for international audience, the argument of biased assessments has not disappeared and reemerges regularly with potentially critical assessments the organization publishes.[29] The question however remains whether the allegations are justified. This leads us to our second hypothesis, which tests whether the OSCE/ODIHR monitoring reports are negatively biased against signatory and affiliated

countries of the Astana Appeal, an umbrella term we use for the Russian-led coalition of post-soviet states questioning the integrity of OSCE election monitoring:

H2: OSCE assessments of elections organized in signatory and supporting countries of Astana Appeal are harsher/more negative than expected.

Data and Methods

To test the aforementioned hypotheses, we analyze an original corpus of 303 OSCE monitoring reports we collected covering the period of 1995 to 2020.[30] It is a mix of elections monitored in different parts of Europe and Eurasia region, with few additions from North America covering western democracies, post-communist countries, and post-conflict societies with different levels of economic development and democratic qualities (see overview in Figure 1). Capitalizing on the advancements of natural language processing and computational linguistics in general, we approach the analysis of textual data from a corpus-based perspective utilizing bag-of-words logic together with a popular *wordscores* scaling algorithm.

The unit of analysis (n = 303) is a final report representing a comprehensive assessment of a specific election made by a monitoring mission. Although not all documents cover all possible dimensions of election monitoring, their general assessment framework is consistent and focuses on whether and to what degree an election or its part met the international standards of free and fair election. Henceforth, we assume that each report we analyze represents a valid approximation of how the OSCE assessed an election in time and space. Apart from standard cleaning and preprocessing[31] we use a pre-trained named entity recognition (NER) model provided by the Allen Institute for AI for extracting any context-specific references to named

entities to avoid a potential location-based bias.[32] The raw corpus after preprocessing consists of 1 528 314 words and 6 584 unique tokens.

[Figure 1 about here]

Using *wordscores* scaling algorithm, the goal is to scale the corpus in order to uncover a latent continuum that defines the overall assessment of the quality of elections, i.e. the extent to which the OSCE considers an election to be *free and fair*. When it comes to underlying logic, *wordscores* algorithm estimates the positions of documents using reference scores for texts whose positions on well-defined a priori dimensions are "known".[33] We combine this approach with a guided bootstrap sampling, a method we propose in order to overcome a problem of selection bias, which inevitably occurs when the reference texts are selected based on close reading (we present a full algorithmic description in Appendix).

As monitoring reports are highly complex, choosing the reference documents is always arbitrary. Moreover, testing has shown that choosing just one pair of documents produces a scale that is not stable and often vary across different pairs. To mitigate this effect we bootstrap the pairs of potentially ten best and ten worst monitoring reports[34] selected based on close reading of collected documents in order to benefit from a good knowledge of the corpus and at the same time to accommodate alternative selection preferences. We choose one election per country potentially covering different nuances of good and bad qualities monitoring reports may focus on in different settings and train 100 *wordscores* models using all combinations of potential pairs in order to stabilize both the scaling scores as well as the standard errors.

The process of bootstrapped scaling gives us stabilized scores (hereafter referred to as the *OSCE election scores)* we can use as an approximation of latent scale of *free and fair* elections.

These scores however needs to be validated extrinsically with proper benchmark. While an objective evaluation of an election is perhaps impossible to make, we decide to use the expert assessments collected by the V-Dem project as a form of empirical yardstick that should tell us how well the selected algorithms perform on a simple scaling task. We use Clean Elections Index (v2xel_frefair) as a standardized score capturing the dimension of free and fair election while covering the whole studied period. The index is not perfect but arguably, it is still superior to any available alternatives in terms of rigor, transparency, methodology, and time span.[35] However, as a robustness check, we provide an additional validation of the scaling results using both the Freedom House and the Polity IV index in the Appendix (Table A2). The results are substantially the same despite the fact that both tested indexes focus on general democratic qualities rather than elections per se.

Although we cannot argue that experts are not exposed to reports under study or do not project their own hidden biases[36], a systematic-level bias in favor of monitoring missions problematizing the whole expert survey is improbable. First, the coders provide their assessment on a highly aggregated level. It means even a source-specific bias is effectively flattened into a number or a code that is an abstraction of much wider range of resources a person with country expertise is exposed through time (e.g. media, research articles, social networks, and monitoring reports). Second, V-Dem's selection criteria for choosing country experts, cross validation of assigned scores and their weighting, and mitigating their biases are thoroughly addressed in the survey's methodology.[37] Third, although our results do not support this argument, we acknowledge that differing information environments across countries might result in an increased reliance of some of the independent experts on findings of OSCE/ODIHR reports. However, this scenario is not prevailing or exclusive. This is most evident under authoritarian regimes in countries such as Russia, Belarus, and Azerbaijan

during elections not observed by the OSCE but still critically evaluated by country experts. There is no indication that the quality of expert assessment significantly deteriorates. Finally, academics and experts themselves occasionally criticize monitoring reports for being biased, inherently recognizing their inconsistencies and problems.[38] This makes us believe the Clean Elections Index, although not perfect, provides a sufficient benchmark we can use for validating the modeled scores.

Empirical Congruence: Monitoring Reports Vs Expert Views

To validate the scaling outcome we compare the OSCE scores with the V-Dem election scores to assess how well the scaling matches the coders' judgement. Figure 2 plots OSCE election scores produced by *wordscores* algorithm against V-Dem election scores. As we can see, there is generally a great deal of consistency between the OSCE and the V-Dem election scores, despite the fact that they employ an entirely different approach to assessing elections (expert surveys vs monitoring missions). Further inspection shows, that rather than geographical clustering, countries are indeed scaled based on qualities of their elections, which empirically covers various political processes like democratization in many Central and Eastern European countries. This is confirmed in Table 1, which reports the relation between the V-Dem scores on the OSCE scores. Model 1, showing the results of a linear regression, confirms this giving us a reasonably strong confidence that the modeled index captures the latent dimension of *free and fair election* quite well.[39]

[Figure 2 about here]

[Table 1 about here]

11

Although the analysis in Model 1 provides strong evidence on existing high congruence between scaled monitoring reports and expert assessments and confirms the expectation that OSCE is a trusted IGO often providing genuine and uncontested assessments[40], we can also see plenty of cases where the scaled scores and expert views disagree. These outliers raise valid questions about whether monitoring reports provide harsher/more lenient assessments for certain contexts than the expert baseline does or it is just noise produced by the scaling algorithms. As the theoretical section suggests, we believe it is the former.

That being said, given the aforementioned methodological differences, it is uncertain and even unlikely whether the relation between V-Dem and the OSCE scores will be linear. This suspicion is confirmed in Model 2, which reports the results of a polynomial regression model. The R² is noticeably higher when compared to the first model, something visually represented in Figure 2.

To analyze the substantive difference between V-Dem and the OSCE scores, we calculate the residuals from Model 2 in Table 1 and use that as an approximation of potential bias. These residuals, the main dependent variable in our analyses, indicate when and to what extent OSCE-election scores present a more positive assessment of an election (positive values) or a more negative one (negative values) than is merited by the V-Dem scores, which thus serves as the benchmark against which to compare the OSCE assessments.[41] This approach takes into account the fact that both scores rely on a different methodology and that some differences are inevitably of an instrumental nature. Our residual-based approach explicitly models the instrumental effects, creating a baseline of a 'normal' (given the methodological differences) relation between the OSCE and V-Dem scores. The values of our dependent variable thus indicate to what degree and in what direction cases deviate from this normal relation.

The first main independent variable we test for is the *year* of election. If the first hypothesis (H1) holds, OSCE reports of more recent elections should be more consistent with the V-Dem scores. For the second hypothesis (H2), the main independent variable is a dummy for being a signatory or supportive state of Astana Appeal and related initiatives *(Astana)*. As the appeal as well as other declarations and initiatives associated with it refer to systematic bias also prior 2004, we expand the relevant window to 1999 when the first Russian attempts to criticize the OSCE activities and profile can be traced. As Ukraine has politically fallen apart with Russia in the recent years, we exclude the country from the group after 2013 (the post Euromaidan era).

Apart from the main independent variables, we further control for two sets of independent variables. First, we focus on contextual factors of election monitoring combining socioeconomic and demographic factors with potential strategic interests of the OSCE monitors (see discussion above). More specifically, we control for *GDP* and *total trade* measured as percentage of GDP as proxies for important markets, volume of *official development assistance* (normalized per capita) as an indicator of dependency on international community, *proportion of population living in urban areas* and *population density* as an indicator of how well the monitors can cover elections in urban and rural areas, and *GDP proportion of total natural resources rents* as an indicator of strategic relevance. Second set of control variables focuses on contextual factors of the monitored elections taking into count mission-specific aspects as well as domestic political climate. We control for a *deployment of full observation mission and its size* (number of observers relative to the size of a country) as an indicator of administrative strength of a mission as well as *regional affiliation of the head of mission* (Western Europe; Eastern Europe; US/Canada) as a proxy for cultural affiliation. Monitored elections are contextualized through variables of *transitional election* as an

indicator of major political change, *turnover election* as an indicator for handover of power, and *legislative election* as contextualizing factor of the race type (see Appendix for an overview of coding rules). Table 2 summarizes the descriptives.[42]

[Table 2 about here]

Analysis

The first hypothesis predicts that the disagreements between the OSCE and V-Dem election scores will decrease over time due to increased competition and an increasingly elaborate monitoring methodology. Table 3 tests this hypothesis.[43] In this analysis, however, we take as dependent variable the absolute value of the residuals. This is because the first hypothesis concerns itself only with the absolute level of bias, regardless of its direction. Therefore, the models test whether the residuals are smaller (closer to zero) as opposed to higher (either under- or over- estimating the quality of an election). This ensures that we are testing whether OSCE scores are more likely to conform to the normal relation with the V-Dem scores if an assessment was made more recently.

Model 1 in Table 3 tests the overall relation between time and OSCE election score bias. Model 2 and 3 gradually make the model more stringent by controlling for the socioeconomic and strategic factors and political factors of the monitored elections. In all three models, we find evidence of a significant and strong decrease in the absolute size of residuals over time (Figure 3 visualizes this trend). In other words, recent OSCE election monitoring reports are more in line with the assessments of experts. This supports hypothesis one, suggesting that efforts made

to elaborate on the methodology of election monitoring pay off and increases the quality of the OSCE election evaluations.

[Table 3 about here]

[Figure 3 about here]

Regarding the control variables, only the distinction between legislative and presidential elections is statistically significant at a p<0.01, with the former showing smaller deviations from the norm than the latter. This can be explained by the nature of legislative races, defined by potentially less tension and more open competition with multiple mandates being contested, as opposed to presidential elections where there is only one winner.[44] As such, presidential elections are thus defined by zero-sum logic with no consolation for second place.[45] The overall more positive assessment of parliamentary elections by both OSCE and expert scores is in line with existing literature which agrees that proportional and multi-mandate elections are less prone to fraud as the incentives for electoral misconduct are lower.[46] Thus, while parliamentary elections can be expected to meet at least some minimal standards of free and fair competition, presidential elections represent much more of a riddle with potential large-scale frauds coming into play.

Because the dependent variable here is the absolute value of the residuals, we are unable to deduce in which direction this bias goes. Therefore, in Table 4, we use the regular values of the residuals. The first model includes only the Astana variable, and Model 2 and Model 3 add the socioeconomic and strategic factors and political factors of the monitored elections. In the first model, the distinction between the Astana and the other countries is not statistically different from zero, but with the addition of the covariates, a difference begins to emerge. In Model 3, we see that the Astana countries are significantly experiencing more positive

evaluations of their elections by the OSCE than merited by V-Dem. Figure 4 visualizes the difference between Astana group and other countries. This is in contrast to what Hypothesis 2 predicted (and Russian-led coalition would hope for).[47]

[Table 4 about here]

[Figure 4 about here]

The conclusions of this finding are twofold. Firstly, the data provides empirical evidence showing that the criticism for alleged negative bias pronounced by Russia and other countries supporting the Astana Appeal is unfounded. This is in line with the majority of literature which has interpreted Russia's criticism as part of the broader effort to delegitimize election observation by OSCE in order to fend off unfavorable assessments and preserve authoritarian regimes fitting to Russia's geopolitical interests in the region.[48] The finding however also points to a second and more important perspective. It shows the bias on the side of OSCE exists but in the opposite direction, meaning that OSCE produced reports that were more positive of the assessed elections than reality warrants. This can be explained by the fact that the Russian-led Astana Appeal and the activities that followed represented a major challenge to the functioning of OSCE/ODIHR to which the institution reacted in various ways.[49] Our data indicates that part of the response that OSCE/ODIHR took in reaction to the raised allegation was to moderate negative assessments of elections in the concerned countries in order to accommodate the Russian-led criticism and avoid proposed structural reforms (see Appendix for an example of lenient assessment towards Russia). In other words, the organization was effectively *pushed* into a more submissive position accommodating a critical voice of powerful actor (or a coalition of actors).

The *pushback effect* Russian-led coalition has successfully imposed on monitoring activities of OSCE/ODIHR can be characterized as a systematic pressure combined with an abuse of structural shortcomings in the functional organization of OSCE as an IGO (e.g. functioning of executive bodies). More generally, the effect is a result of a political pressure leading to a change of position that is seen as unwanted or less preferable. Although different forms of bias might come from external pressure, the *pushback effect* captures a specific dynamics that is long-term, political in nature, and focuses on changing the core principles of election monitoring.

Apart from the pushback effect, it is important to acknowledge that at least in some cases, an additional contributing factor might play a role in explaining the observed leniency. As a number of authoritarian regimes regularly alters their strategies for election manipulation, the capacity of election monitors to verify and document them can be regularly challenged as well. If international monitors are not able to keep up with the advancements of election manipulation, it also might lead to a more positive assessment than reality warrants.[50]

Regarding the control variables, we see that higher GDP, GDP proportion of total natural resources rents, official development assistance, and legislative elections are associated with more positive assessments. These variables tell three distinct stories. First, there is a positive bias towards strategic markets either in terms of mere size (GDP) or their importance (natural resources) which are in line with criticism that economic interests might interfere with international organizations' monitoring goals.[51] Second, positive bias towards recipients of ODA appear to be in line with the argument that IGO member states may attach particular importance to countries that receive more foreign aid and treat these more leniently as a reflection of their commitment.[52] Lastly, the positive assessment towards parliamentary elections can be explained by a relatively higher level of competitiveness (i.e. more seats to

compete for) than we observe in zero-sum contests such as presidential elections with no consolation for attaining even one vote fewer than one's rival.[53] Only one variable shows significant and negative effect (i.e. negative bias) – the size of observation mission. We theorize that it mostly reflects a relative number of observers who carry out the observations. In this context, more observers being deployed can spot more irregularities, hence provide on average a more critical assessment. Moreover, bigger monitoring missions are probably more often allowed in countries where the OSCE assessment can be critical. This is in line with the practice of monitoring of elections in countries with authoritarian or repressive governments where monitoring missions are more often than not smaller than the size of a country would require (e.g. Russia).

Conclusion

The paper analyzes a quarter of century of election monitoring in Europe. Based on the original corpus of OSCE monitoring reports, we explore the existing biases in raw textual data and assess them against independent expert views. Our results show that OSCE is highly consistent with the expert opinions in assessing whether and to what degree an election can be considered free and fair. In this context, the OSCE conducts the overwhelming majority of its assessments with a high degree of professional integrity and continues to improve the quality of its work with time (H1). However, our analysis has also identified several biases of which the most relevant refers to a positive bias towards Russia and its allies effectively showing a defensive position the OSCE has when it comes to election monitoring in these countries (H2).

The paper should be read as a test of existing theoretical arguments using original data with an aspiration to understand how political bias works in predominantly (but not exclusively)

European context. As we showed in the previous section, the story of bias is both political and strategic. On the one hand, the congruence of reports and expert assessments increases with time, showing the overall standards of election monitoring being continuously improved. On the other, we show that some contexts are more prone to be assessed with a bias than others. Important markets as well as countries with strategic interests get more lenient assessments, which problematizes the legacy of OSCE election monitoring whose impartiality seemingly reaches its limits when it comes to economic and political realities of the OSCE region. On the other hand, we show that the accusation of double standards expressed by the Russian-led coalition of states indeed exists, but in the opposite direction than pictured by the concerned governments. It indicates a defensive position the OSCE was *pushed* to over the years of political squabbles, which apparently helped Russia and its allies to receive more moderate assessments than reality warrants. We conceptualize this mechanism as a *pushback effect* which explains the lenient assessments as a result of systematic pressure imposed by Russian-led coalition on OSCE as an IGO.

Overall, the paper presents a complex picture of the OSCE's legacy of election observation missions in Europe in the past quarter of century. Despite the declared high standards, OSCE/ODIHR has not always delivered on the principles of impartiality and accuracy of assessments when confronted with vested interests of OSCE member states and complex geopolitical realities of the OSCE area. Although it is not surprising that an international organization composed of national governments yields to political pressures and concealed national interests, the existence of bias is not justifiable considering practical implications that election assessments have for domestic and international audiences. We believe that the evidence we have presented here has given us more insight into the workings of international

monitoring organizations, the output they produce, the political goals they seek to balance, and

pressure they might face.

Notes

1   Hyde and Marinov, "Information and Self-Enforcing Democracy: The Role of International Election Observation"; Norris, *Why Electoral Integrity Matters*.
2   Daxecker, "The Cost of Exposing Cheating"; Simpser and Donno, "Can International Election Monitoring Harm Governance?," July 19, 2012; Kohnert, "Election Observation in Nigeria and Madagascar: Diplomatic vs. Technocratic Bias."
3   Kelley, "D-Minus Elections: The Politics and Norms of International Election Observation"; Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*.
4   Kavakli and Kuhn, "Dangerous Contenders: Election Monitors, Islamic Opposition Parties and Terrorism"; Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*; see potential examples of such assessments in the Appendix; Kelley, "D-Minus Elections: The Politics and Norms of International Election Observation."
5   Carothers, "The Rise of Election Monitoring: The Observers Observed."
6   United Nations, "Declaration of Principles for International Election Observation and Code of Conduct for International Election Observers."
7   Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*; Kelley, "D-Minus Elections: The Politics and Norms of International Election Observation."
8   Kavakli and Kuhn, "Dangerous Contenders: Election Monitors, Islamic Opposition Parties and Terrorism."
9   Elklit and Reynolds, "A Framework for the Systematic Study of Election Quality."; Elklit and Svensson, "What Makes Elections Free and Fair?"; EODS, *Compendium of International Standards for Elections*.
10 Regalia, "Working for Democracy: The Effectiveness of Election Observation"; Bader and Schmeets, "Is International Election Observation Credible? Evidence from Organization for Security and Co-Operation in Europe Missions"; Hyde, "The Observer Effect in International Politics: Evidence from a Natural Experiment."
11 van Peski and Schmeets, "The 'C-Factor': Impact of The Nationality of Observers on Observing Elections."
12 Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space"; Fawn, *International Organizations and Internal Conditionality: Making Norms Matter*.
13 Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*.
14 United Nations, "Declaration of Principles for International Election Observation and Code of Conduct for International Election Observers."
15 OSCE/ODIHR, *Election Observation Handbook: Sixth Edition*; EODS, *Handbook for European Union Election Observation*.

16 Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*.

17 Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space"; Sikkink, "Restructuring World Politics: The Limits and Assymmetries of Soft Power"; Norris, "Does the World Agree About Standards of Electoral Integrity? Evidence for the Diffusion of Global Norms."

18 Eicher, "Improving OSCE Election Observation"; Laanela, "Beyond the Checklist: Addressing New Challenges in Election Observation Methodology."

19 Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*.

20 cf. Roth, "Despots Masquerading as Democrats: World Report 2008"; Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space"; Bjornlund, *Beyond Free and Fair: Monitoring Elections and Building Democracy*; Kelley, "Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring."

21 van Willigen, "Effective Multilateralism between Unequal Partners: The EU in the OSCE"; Kropatcheva, "Russia and the Role of the OSCE in European Security: A 'Forum' for Dialog or a 'Battlefield' of Interests?"; van Willigen and Koops, "The EU's Relationship with NATO and OSCE"; Marciacq and Jaramillo, "When the European Union Speaks on Behalf of Non-European Union States: A Critical Appraisal of the European Union's Alignment Mechanism in Multilateral Fora."

22 Since ODIHR's establishment in 1994, every director of the institution has been selected from NATO or EU member states. Likewise, the composition of EOM Core Teams (in particular those dealing with assessment of elections) is dominated by nationals of NATO or EU.

23 Novosad and Werker, "Who Runs the International System? Power and the Staffing of the United Nations Secretariat"; Murdoch et al., "Do International Institutions Matter? Socialization and International Bureaucrats."

24 Orenstein, *The Lands in Between: Russia Vs. The West and the New Politics of Hybrid War*; Marten, "Reconsidering NATO Expansion: A Counterfactual Analysis of Russia and the West in the 1990s."

25 Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space."

26 Fawn, *International Organizations and Internal Conditionality: Making Norms Matter*, 65–66; cf. Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space," 1138.

27 Fawn, *International Organizations and Internal Conditionality: Making Norms Matter*, 66.

28 Siegel, "The Issue of the Office for Democratic Institutions and Human Rights," 143; Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space," 1141.

29 Evers, "OSCE Election Observation: Commitments, Methodology, Criticism."

30 The corpus includes final reports available on the official OSCE website as of May 1, 2021: 192 parliamentary elections, 15 general elections, 89 presidential elections, and 7 elections combining parliamentary and presidential elections. The corpus covers full reports from both the Election Assessment Missions and the Election Observation Missions. We control for a potential difference using the *mission size* variable.

31 Each document is manually cleaned off of its title page, content, appendix, footers, and headers. After lemmatization of raw text, documents are cleaned off of stop-words, infrequent words (occurring less than 5 times across corpus), numbers, punctuation, and whitespaces. All upper case characters are transformed to lower case. Pre-processing and analysis is done in R package *quanteda*. Benoit et al., "Quanteda: An R Package for the Quantitative Analysis of Textual Data.".

32 AllenNLP, "Named Entity Recognition."

33 Laver, Benoit, and Garry, "Extracting Policy Positions from Political Texts Using Words as Data."

34 We arbitrary assign a value of ten [10] to a reference document that could be chosen as the potentially most positive assessment and a value of one [1] to the potentially most negative assessment in our corpus. See Table A1 in the appendix for the specific monitoring reports identified as potentially most praising/critical.

35 Coppedge, Gerring, and Lindberg, "Varieties of Democracy"; Coppedge et al., "V-Dem Comparisons and Contrasts with Other Measurement Projects"; Coppedge et al., "Measuring High Level Democratic Principles Using the V-Dem Data"; Freedom House, "Freedom House"; Polity IV Project, "Polity IV."

36 Martínez i Coma and van Ham, "Can Experts Judge Elections? Testing the Validity of Expert Judgments for Measuring Election Integrity."

37 see a lengthy discussion on selection procedure and profiles of country coders in Coppedge et al., "The Methodology of 'Varieties of Democracy' (V-Dem)."

38 Kohnert, "Election Observation in Nigeria and Madagascar: Diplomatic vs. Technocratic Bias"; Mendelson, "Democracy Assistance and Political Transition in Russia: Between Success and Failure," 104; Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space," 1136–38.

39 Given the seemingly non-linear relation between the OSCE and V-Dem election scores, we tried varies alternative specification of the regression model (e.g. log transforming the OSCE scores), but the base model showed the greatest fit with the data.

40 Kelley, *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*.

41 To be sure, we checked whether alternatives to the V-Dem scores yield the same results by repeating the analyses, but with the Freedom House score. The results, reported in the Appendix, are substantively the same (see Table A3 and Table A4).

42 In order to avoid confounding effects, we do not include the variable 'Astana' in the models of Table 3. As a robustness check, we repeat the analysis with it in Table A5 (Appendix).

43 The models in Table 3 as well as Table 4 account for the fact that the observations are not independent but can come from the same country, through clustered standard errors.

44 Dawson, "Electoral Fraud and the Paradox of Political Competition."

45 Linz, "Transitions to Democracy."

46 Birch, "Electoral Systems and Electoral Misconduct"; Lehoucq and Kolev, "Varying the Un-Variable: Social Structure, Electoral Formulae, and Election Quality"; Ruiz-Rufino, "When Do Electoral Institutions Trigger Electoral Misconduct?"

47 For the same as we excluded the variable 'Astana' in Table 3, we leave out the variable 'Year' in the models of Table 4. Here too, however, we ran a robustness check, which did include it as a covariate. The model reported in Table A6 supports the results shown here. As another robustness check, Table A7 in Appendix presents the models with Astana variable using period after 2003 instead of 1999.

48 Evers, "OSCE Election Observation: Commitments, Methodology, Criticism"; Zellner, "Russia and the OSCE: From High Hopes to Disillusionment"; Ghebali, "Growing Pains at the OSCE: The Rise and Fall of Russia's Pan-European Expectations."
49 Fawn, "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space."
50 Simpser and Donno, "Can International Election Monitoring Harm Governance?," 2012; Schedler, "The Menu of Manipulation."
51 Kelley, "D-Minus Elections: The Politics and Norms of International Election Observation"; Ross, "Oil and Gas Data, 1932-2011."
52 Stone, "The Scope of IMF Conditionality Randall"; Kelley, "D-Minus Elections: The Politics and Norms of International Election Observation"; Hynek and Marton, *Statebuilding in Afghanistan: Multinational Contributions to Reconstruction*.
53 Dawson, "Electoral Fraud and the Paradox of Political Competition."

References

AllenNLP. "Named Entity Recognition," 2020. https://demo.allennlp.org/named-entity-recognition.

Bader, Max, and Hans Schmeets. "Is International Election Observation Credible? Evidence from Organization for Security and Co-Operation in Europe Missions." *Research and Politics* 1, no. 2 (2014): 1–6.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3, no. 30 (2018): 774.

Birch, Sarah. "Electoral Systems and Electoral Misconduct." *Comparative Political Studies* 40, no. 12 (2007): 1533–56.

Bjornlund, Eric. *Beyond Free and Fair: Monitoring Elections and Building Democracy*. Washington D.C.: Woodrow Wilson Center Press; Johns Hopkins University Press, 2004.

Carothers, Thomas. "The Rise of Election Monitoring: The Observers Observed." *Journal of Democracy* 8, no. 3 (1997): 17–31.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Joshua Krusell, Juraj Medzihorsky, Josefine Pernes, Svend-Erik Skaaning, et al. "The Methodology of 'Varieties of Democracy' (V-Dem)." *Bulletin of Sociological Methodology* 143, no. 1 (2019): 107–33.

Coppedge, Michael, John Gerring, and Staffan I. Lindberg. "Varieties of Democracy," 2017. https://www.v-dem.net/en/.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, and Jan Teorell. "V-Dem Comparisons and Contrasts with Other Measurement Projects." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 2017.

Coppedge, Michael, Staffan Lindberg, Svend Erik Skaaning, and Jan Teorell. "Measuring High Level Democratic Principles Using the V-Dem Data." *International Political Science Review* 37, no. 5 (2015): 580–93.

Dawson, Stephen. "Electoral Fraud and the Paradox of Political Competition." *Journal of Elections, Public Opinion and Parties* Online (2020): 1–20.

Daxecker, Ursula E. "The Cost of Exposing Cheating." *Journal of Peace Research* 49, no. 4 (July 10, 2012): 503–16.

Eicher, Peter. "Improving OSCE Election Observation." *Security and Human Rights Human*

Rights 20, no. 4 (2009): 264–76.

Elklit, Jørgen, and Andrew Reynolds. "A Framework for the Systematic Study of Election Quality." *Democratization* 12, no. 2 (2005): 147–62.

Elklit, Jørgen, and Palle Svensson. "What Makes Elections Free and Fair?" *Journal of Democracy* 8, no. 3 (1997): 32–46.

EODS. *Compendium of International Standards for Elections*. Brussels: Publication Office of the European Union, 2016.

———. *Handbook for European Union Election Observation*. Luxembourg: Publication Office of the European Union, 2016.

Evers, Frank. "OSCE Election Observation: Commitments, Methodology, Criticism." *OSCE Yearbook 2009*, 2010.

Fawn, Rick. "Battle over the Box: International Election Observation Missions , Political Competition and Retrenchment in the Post-Soviet Space." *International Affair* 82, no. 6 (2006): 1133–53.

———. *International Organizations and Internal Conditionality: Making Norms Matter*. New York: Palgrave Macmillan, 2013.

Freedom House. "Freedom House." Washington, 2017. https://freedomhouse.org/.

Ghebali, Victor-Yves. "Growing Pains at the OSCE: The Rise and Fall of Russia's Pan-European Expectations." *Cambridge Review of International Affairs* 18, no. 3 (2005): 375–88.

Hyde, Susan D. "The Observer Effect in International Politics: Evidence from a Natural Experiment." *World Politics* 60, no. 01 (2007): 37–63.

Hyde, Susan D., and Nikolay Marinov. "Information and Self-Enforcing Democracy: The Role of International Election Observation." *International Organization* 68, no. 02 (2014): 329–59.

Hynek, Nik, and Péter Marton. *Statebuilding in Afghanistan: Multinational Contributions to Reconstruction*. London and New York: Routledge, 2012.

Kavakli, Kerim, and Patrick Kuhn. "Dangerous Contenders: Election Monitors, Islamic Opposition Parties and Terrorism." *International Organization* 74, no. 1 (2020): 145–64.

Kelley, Judith G. "Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring." *International Organization* 62, no. 02 (2008): 221–55.

———. "D-Minus Elections: The Politics and Norms of International Election Observation." *International Organization* 63, no. 04 (2009): 765–87.

———. *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*. Princeton: Princeton University Press, 2012.

Kohnert, Dirk. "Election Observation in Nigeria and Madagascar: Diplomatic vs. Technocratic Bias." *Review of African Political Economy* 31, no. 99 (2004): 83–101.

Kropatcheva, Elena. "Russia and the Role of the OSCE in European Security: A 'Forum' for Dialog or a 'Battlefield' of Interests?" *European Security* 21, no. 3 (2012): 370–94.

Laanela, Therese Pearce. "Beyond the Checklist: Addressing New Challenges in Election Observation Methodology." *Nordic Journal of Human Rights* 35, no. 4 (2017): 327–40.

Laver, Michael, Kenneth Benoit, and John Garry. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97, no. 2 (2003): 311–31.

Lehoucq, Fabrice, and Kiril Kolev. "Varying the Un-Variable: Social Structure, Electoral Formulae, and Election Quality." *Political Research Quarterly* 68, no. 2 (2015): 240–52.

Linz, Juan J. "Transitions to Democracy." *The Washington Quarterly* 13, no. 3 (1990): 143–64.

Marciacq, Florent, and Natalia Sanmartín Jaramillo. "When the European Union Speaks on

Behalf of Non-European Union States: A Critical Appraisal of the European Union's Alignment Mechanism in Multilateral Fora." *European Security* 24, no. 2 (2015): 203–20.

Marten, Kimberly. "Reconsidering NATO Expansion: A Counterfactual Analysis of Russia and the West in the 1990s." *European Journal of International Security* 3, no. 2 (2018): 135–61.

Martínez i Coma, Ferran, and Carolien van Ham. "Can Experts Judge Elections? Testing the Validity of Expert Judgments for Measuring Election Integrity." *European Journal of Political Research* 54, no. 2 (2015): 305–25.

Mendelson, Sarah E. "Democracy Assistance and Political Transition in Russia: Between Success and Failure." *International Security* 25, no. 4 (2001): 68–106.

Murdoch, Zuzana, Hussein Kassim, Sara Connolly, and Benny Geys. "Do International Institutions Matter? Socialization and International Bureaucrats." *European Journal of International Relations* 25, no. 3 (2019): 852–77.

Norris, Pippa. "Does the World Agree About Standards of Electoral Integrity? Evidence for the Diffusion of Global Norms." *Electoral Studies* 32, no. 4 (2013): 576–88.

———. *Why Electoral Integrity Matters*. New York: Cambridge University Press, 2014.

Novosad, Paul, and Eric Werker. "Who Runs the International System? Power and the Staffing of the United Nations Secretariat." *The Review of International Organizations* 14, no. 1 (2019): 1–33.

Orenstein, Mitchell A. *The Lands in Between: Russia Vs. The West and the New Politics of Hybrid War*. Oxford: Oxford University Press, 2019.

OSCE/ODIHR. *Election Observation Handbook: Sixth Edition*. Warsaw: OSCE/ODIHR, 2010.

Peski, Caecilia van, and Hans Schmeets. "The 'C-Factor': Impact of The Nationality of Observers on Observing Elections." In *International Election Observation and Assessment of Elections*, edited by Hans Schmeets. The Hague: Statistics Netherlands, 2011.

Polity IV Project. "Polity IV," 2012. http://www.systemicpeace.org/polity/polity4.htm.

Regalia, Marta. "Working for Democracy: The Effectiveness of Election Observation." In *International Election Observation and Assessment of Elections*, edited by Hans Schmeets. The Hague: Statistics Netherlands, 2011.

Ross, Michael L. "Oil and Gas Data, 1932-2011," 2013. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DENAF5.

Roth, Kenneth. "Despots Masquerading as Democrats: World Report 2008." New York, 2008.

Ruiz-Rufino, Rubén. "When Do Electoral Institutions Trigger Electoral Misconduct?" *Democratization* 25, no. 2 (2018): 331–50.

Schedler, Andreas. "The Menu of Manipulation." *Journal of Democracy* 13, no. 2 (2002): 36–50.

Siegel, Walter. "The Issue of the Office for Democratic Institutions and Human Rights." In *The OSCE at a Turning Point*, edited by Daniel Warner. Geneva: Graduate Institute of International Studies, 2007.

Sikkink, Kathryn. "Restructuring World Politics: The Limits and Assymmetries of Soft Power." In *Restructuring World Politics: Transnational Social Movements, Networks, and Norms*, edited by S Khagram, James Riker, and Kathryn Sikkink. Minneapolis: University of Minnesota Press, 2002.

Simpser, Alberto, and Daniela Donno. "Can International Election Monitoring Harm Governance?" *Journal of Politics* 74, no. 2 (2012): 501–13.

———. "Can International Election Monitoring Harm Governance?" *Journa of Politics* 74,

no. 2 (July 19, 2012): 501–13.

Stone, Randall W. "The Scope of IMF Conditionality Randall." *International Organization* 62, no. 4 (2008): 589–620.

United Nations. "Declaration of Principles for International Election Observation and Code of Conduct for International Election Observers," 2005, 15.

Willigen, Niels van. "Effective Multilateralism between Unequal Partners: The EU in the OSCE." In *The EU and Effective Multilateralism*, edited by Edith Drieskens and Louise G. van Schaik, 154–70. Routledge, 2014.

Willigen, Niels van, and Joachim A. Koops. "The EU's Relationship with NATO and OSCE." *The SAGE Handbook of European Foreign Policy* 2 (2015): 736–49.

Zellner, Wolfgang. "Russia and the OSCE: From High Hopes to Disillusionment." *Cambridge Review of International Affairs* 18, no. 3 (2005): 389–402.

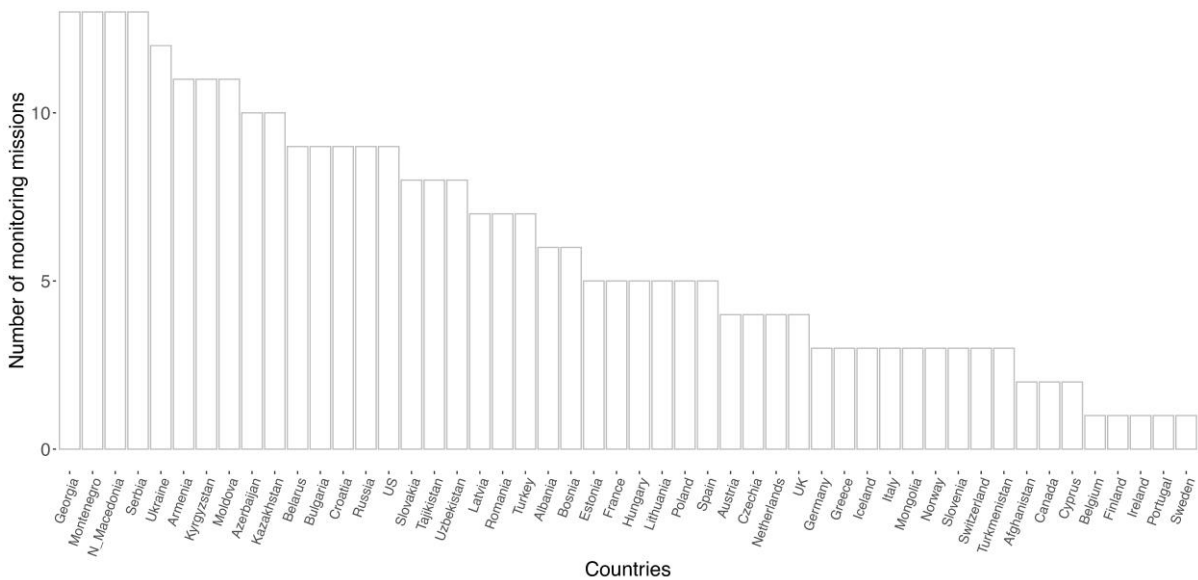Figure 1. Number of monitored elections per country

Figure 2. OSCE and V-Dem election scores
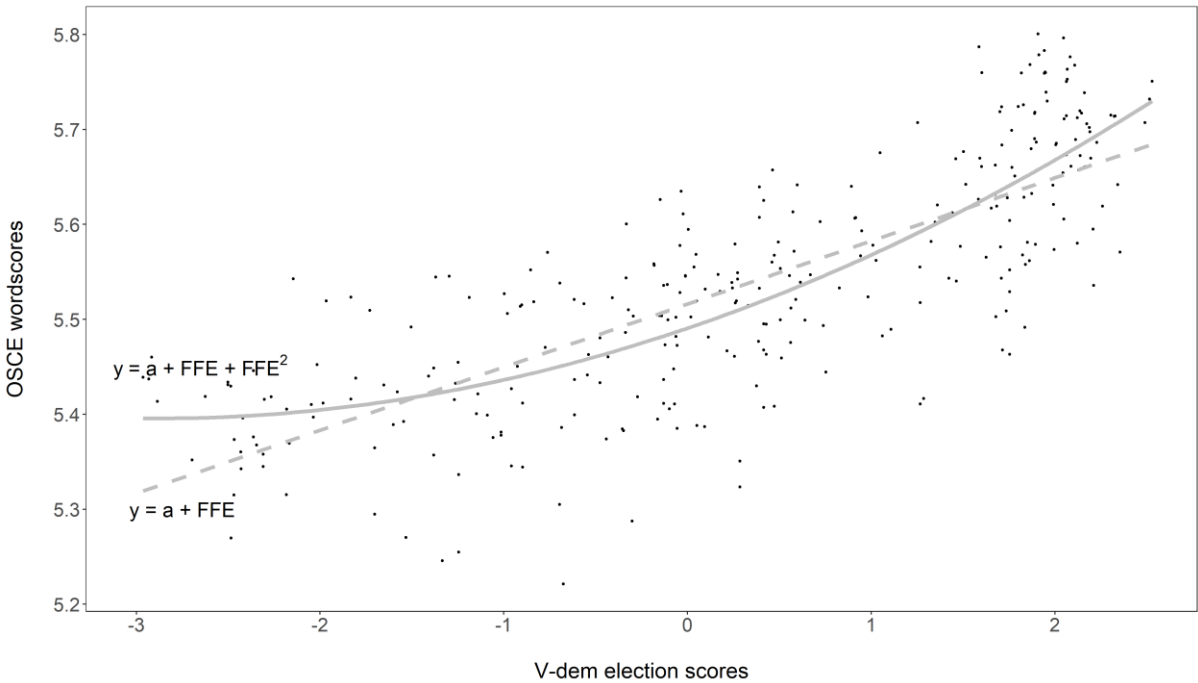
Figure 3. Decrease in the absolute size of residuals over time
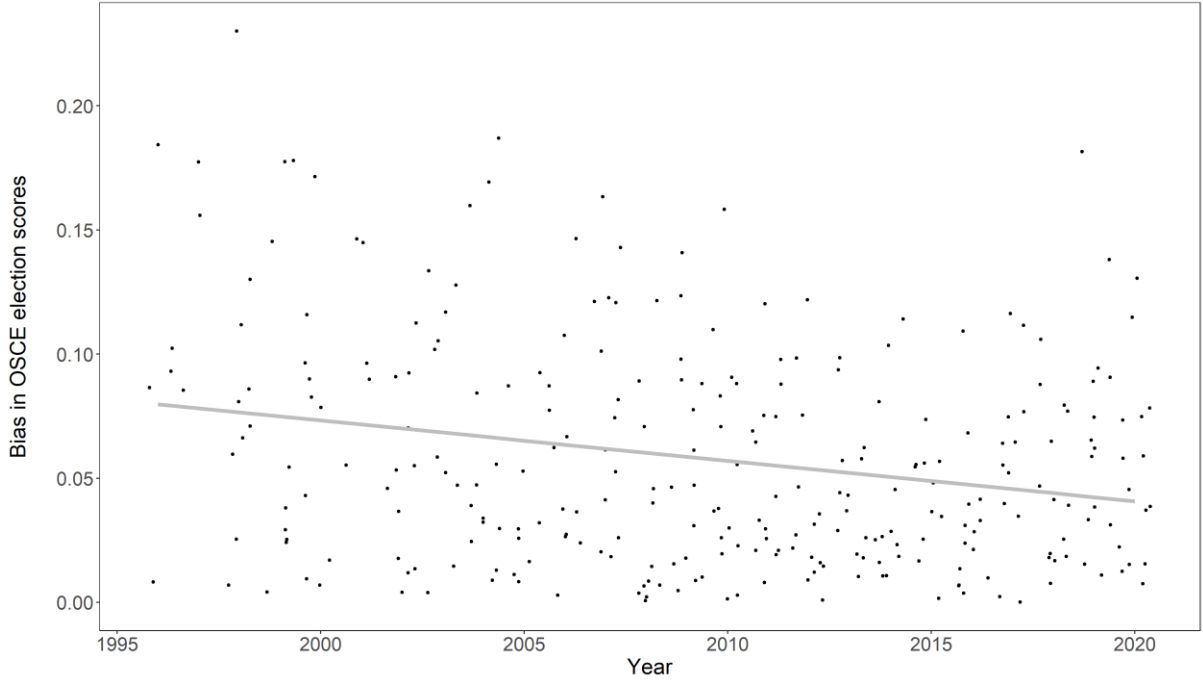
Figure 4. Difference between Astana group and other countries.
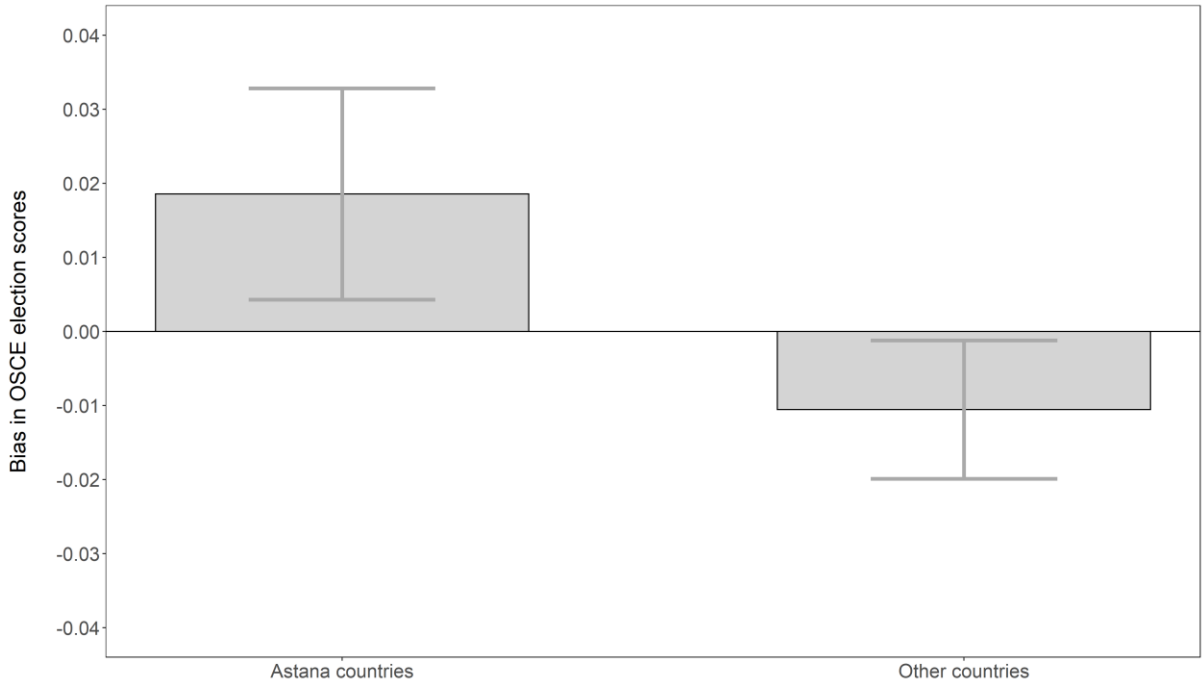
Table 1. The relation between OSCE and V-Dem election scores

	Model 1: Linear model			Model 2: Polynomial model		
	B	S.E.	Sig.	B	S.E.	Sig.
FFE	0.07	0.00	***	0.07	0.00	***
FFE²				0.01	0.00	***
Constant	5.52	0.00	***	5.49	0.01	***
n	303			303		
Adj. R²	60.96%			64.46%		

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table 2. Overview of the descriptives

	Mean	SD	Min.	Max.
Residuals	0.00	0.07	-0.23	0.16
|Residuals|	0.06	0.05	0.00	0.23
Year	2009.61	6.79	1995	2020
Astana	0.27	0.45	0.00	1.00
GDP	14434.55	17761.75	415.46	91549.04
Total trade (% of GDP)	91.58	33.47	22.15	184.48
Official development assistance	41.80	60.04	0.00	278.08
Proportion of population living in urban areas (%)	61.52	14.01	23.53	97.51
Population density	82.52	68.93	1.85	508.80
GDP proportion of total natural resources rents (%)	4.39	8.03	0.00	41.21
Deployment of full observation mission and its size	2.39	1.29	1.00	5.00
Mission head WE	0.56	0.50	0.00	1.00
Mission head EE	0.30	0.46	0.00	1.00
Mission head US/CA	0.14	0.35	0.00	1.00
Transitional election	0.05	0.21	0.00	1.00
Turnover election	0.34	0.47	0.00	1.00
Legislative election	0.71	0.46	0.00	1.00

Table 3. Analyses of the absolute bias in OSCE election scores

	Model 1			Model 2			Model 3		
	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.
Year	-1.63E-03	3.86E-04	***	-1.67E-03	4.03E-04	***	-1.52E-03	4.08E-04	***
GDP				0	0		0	0	
Total trade (% of GDP)				0	0		0	0	
Proportion of population living in urban areas				0	0		0	0	
Population density				0	0		0	0	
GDP proportion of total natural resources rents				0	0		0	0	
Deployment of full observation mission and its size							0	0	
Official development assistance							0	0	
Mission head WE (ref. cat.)									
Mission head EE							0	0.01	
Mission head US/CA							0.01	0.01	
Transitional election							0.02	0.01	
Turnover election							-0.01	0.01	
Legislative election							-1.81E-02	5.95E-03	**
Intercept	3.33	0.77	***	3.43	0.81	***	3.114	0.8188	***
n	288			288			288		
Adj R²	5.53%			5.37%			8.63%		

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table 4. Analyses of the bias in OSCE election scores

	Model 1			Model 2			Model 3		
	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.
Astana	0.00	0.01		1.80E-02	1.04E-02	†	2.91E-02	9.76E-03	**
GDP				1.69E-06	3.24E-07	***	1.33E-06	3.09E-07	***
Total trade (% of GDP)				0.00	0.00		0.00	0.00	
Proportion of population living in urban areas				-8.49E-04	3.78E-04	*	0.00	0.00	†
Population density				0.00	0.00		0.00	0.00	
GDP proportion of total natural resources rents				0.00	0.00		9.75E-04	4.82E-04	*
Deployment of full observation mission and its size							-8.43E-03	3.64E-03	*
Official development assistance							2.50E-04	7.27E-05	***
Mission head WE (ref. cat.)									
Mission head EE							0.01	0.01	
Mission head US/CA							-0.02	0.01	†
Transitional election							-0.02	0.02	
Turnover election							-0.02	0.01	†
Legislative election							7.65E-02	8.19E-03	***
Intercept	0.00	0.01		0.00	0.03		-0.04	0.03	†
n	288			288			288		
Adj R²	0.07%			8.95%			34.59%		

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A1. Reports used for bootstrapping the output of *wordscores* algorithm

Positive Reports	Negative Report
2007 Parliamentary Election in Belgium	1997 Parliamentary Election in Albania
2011 Parliamentary Election in Estonia	2003 Presidential Election in Armenia
2011 Parliamentary Election in Finland	2018 Presidential Election in Azerbaijan
2007 Presidential Election in France	2004 Parliamentary Election in Belarus
2009 Parliamentary Election in Germany	2003 Parliamentary Election in Georgia
2009 Parliamentary Election in Greece	2000 Presidential Election in Kyrgyzstan
2007 Parliamentary Election in Ireland	2020 Presidential Election in Tajikistan
2018 Parliamentary Election in Italy	2017 Presidential Election in Turkmenistan
2009 Parliamentary Election in Portugal	2004 Presidential Election in Ukraine
2008 Parliamentary Election in Spain	1999 Parliamentary Election in Uzbekistan

Table A2. Correlation analysis of OSCE word scores, Polity IV, V-Dem, and Freedom House

	OSCE word scores	V-Dem	Freedom House	Polity IV
OSCE word scores	1	0.78***	-0.74***	0.65***
V-Dem		1	-0.91***	0.83***
Freedom House			1	-0.88***
Polity IV				1

Note: Pearson's product-moment correlation; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A3. Analyses of the absolute bias in OSCE election scores, with Freedom House Scores

	Model 1			Model 2			Model 3		
	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.
Year	-1.14E-03	3.84E-04	**	-1.08E-03	4.03E-04	**	-8.84E-04	4.07E-04	*
Astana				-0.01	0.01		-0.01	0.01	
GDP				0.00	0.00		0.00	0.00	
Total trade				0.00	0.00		0.00	0.00	
Proportion of population living in urban areas				0.00	0.00		0.00	0.00	
Population density				0.00	0.00		0.00	0.00	
GDP proportion of total natural resources rents				0.00	0.00		0.00	0.00	
Deployment of full observation mission and its size							0.00	0.00	
Official development assistance							0.00	0.00	
Mission head WE (ref. cat.)									
Mission head EE							0.00	0.01	
Mission head US/CA							0.01	0.01	†
Transitional election							0.02	0.01	†
Turnover election							-0.01	0.01	†
Legislative election							-1.60E-02	5.91E-03	**
Intercept	2.34	0.77	**	2.23	0.81	**	1.84	0.82	*
n		288			288			288	
Adj R²		2.64%			1.84%			6.27%	

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A4. Analyses of the bias in OSCE election scores, with Freedom House Scores

	Model 1			Model 2			Model 3		
	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.
Astana	0.00	0.01		0.01	0.01		2.04E-02	9.93E-03	*
GDP				9.16E-07	3.34E-07	**	0.00	0.00	†
Total trade (% of GDP)				0.00	0.00		0.00	0.00	
Proportion of population living in urban areas				-8.93E-04	3.90E-04	*	-7.89E-04	3.44E-04	*
Population density				0.00	0.00	†	1.23E-04	5.88E-05	*
GDP proportion of total natural resources rents				0.00	0.00		0.00	0.00	
Deployment of full observation mission and its size							-0.01	0.00	
Official development assistance							2.00E-04	7.40E-05	**
Mission head WE (ref. cat.)									
Mission head EE							0.00	0.01	
Mission head US/CA							-0.01	0.01	
Transitional election							-0.04	0.02	†
Turnover election							-0.01	0.01	
Legislative election							8.65E-02	8.33E-03	***
Intercept	0.00	0.01		0.03	0.03		-0.02	0.03	
n		288			288			288	
Adj R²		0.00%			2.84%			32.13%	

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A5: Analyses of the absolute bias in OSCE election scores, controlling for Astana membership

	Model 1			Model 2			Model 3		
	B	S.E.	Sig.	B	S.E.	Sig.	B	S.E.	Sig.
Year	-1.63E-03	3.86E-04	***	-1.66E-03	4.04E-04	***	-1.48E-03	4.10E-04	***
Astana				0.00	0.01		-0.01	0.01	
GDP				0.00	0.00		0.00	0.00	
Total trade (% of GDP)				0.00	0.00		0.00	0.00	
Proportion of population living in urban areas				0.00	0.00		0.00	0.00	
Population density				0.00	0.00		0.00	0.00	
GDP proportion of total natural resources rents				0.00	0.00		0.00	0.00	
Deployment of full observation mission and its size							0.00	0.00	
Official development assistance							0.00	0.00	
Mission head WE (ref. cat.)									
Mission head EE							0.00	0.01	
Mission head US/CA							1.33E-02	7.93E-03	†
Transitional election							0.02	0.01	
Turnover election							-0.01	0.01	
Legislative election							-1.83E-02	5.95E-03	**
Intercept	3.33	0.77	***	3.40	0.81	***	3.05	0.82	***
n		288			288			288	
Adj R²		5.53%			5.19%			8.57%	

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A6: Analyses of the bias in OSCE election scores, controlling for year

	Model 1		
	B	S.E.	Sig.
Astana	2.52E-02	9.50E-03	**
GDP	1.07E-06	3.06E-07	***
Total trade (% of GDP)	0.00	0.00	
Proportion of population living in urban areas	-6.46E-04	3.27E-04	*
Population density	0.00	0.00	†
GDP proportion of total natural resources rents	0.00	0.00	†
Deployment of full observation mission and its size	-7.56E-03	3.53E-03	*
Official development assistance	2.21E-04	7.08E-05	**
Mission head WE (ref. cat.)			
Mission head EE	0.01	0.01	
Mission head US/CA	-0.01	0.01	
Transitional election	-0.02	0.02	
Turnover election	-1.72E-02	8.16E-03	*
Legislative election	7.61E-02	7.94E-03	***
Year	2.36E-03	5.47E-04	***
Intercept	-4.78	1.10	***
n	288		
Adj R²	38.55%		

Note: Clustered OLS regression; †p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A7: Analyses of the bias in OSCE election scores in elections after 2003

	Model 1		
	B	S.E.	Sig.
Astana	2.86E-02	1.01E-02	**
GDP	8.27E-07	2.95E-07	**
Total trade (% of GDP)	-6.46E-05	1.04E-04	
Proportion of population living in urban areas	0.00	0.00	†
Population density	1.24E-04	5.11E-05	*
GDP proportion of total natural resources rents	0.00	0.00	
Deployment of full observation mission and its size	-1.03E-02	3.41E-03	**
Official development assistance	2.84E-04	7.13E-05	***
Mission head WE (ref. cat.)			
Mission head EE	0.00	0.01	
Mission head US/CA	-0.01	0.01	
Transitional election	-0.01	0.02	
Turnover election	-2.70E-02	8.02E-03	***
Legislative election	7.26E-02	8.10E-03	***
Intercept	0.00	0.03	
n	288		
Adj R²	37.85%		

Note: Clustered OLS regression; †$p<0.1$; *$p<0.05$; **$p<0.01$; ***$p<0.001$

Bootstrapping wordscores models

The idea behind bootstrapping wordscores models comes from a need to select reference documents that are used for scaling virgin documents (i.e. documents with unknown scores). More often than not, the scores of known documents are assigned arbitrarily projecting different kinds of selection bias. If a wordscores model is run with different sets of arbitrarily similar documents, the assigned scores logically differ. Even two documents that are very similar in nature will provide a slightly different results coming from different vocabulary both documents contain. Although the implementation of wordscores algorithm in *quanteda* package[1] allows using multiple documents for estimating positions of virgin documents, this setting is not always suitable for corpus with complex issue structure. As wordscores algorithm and its properties is discussed extensively elsewhere[2], this short overview focuses solely on methodological innovation introduced in the article.

Rationale

The proposed approach comes from a real world research problem when a researcher needs to pick a pair or a set of documents for wordscores scaling. In the case of OSCE corpus, that means selecting reports that assess elections with very poor standards on the one hand, and those which are conducted in accordance with international standards on the other. With 303 documents, it is hard to pick just one document from each group as multiple elections fulfil these preliminary criteria. The same goes for different groupings with multiple documents. Testing has shown that choosing different groups/pairs of documents produces different results highlighting the issue of selection bias. Moreover, grouping documents which are similar in kind but not identical might overlook the unique features each report provides as a scaling seed. To address this problem, the paper proposes an approach based on selection of potential references documents and rotating them in an analytical pipeline using bootstrapping logic. For this purpose, the paper identifies ten monitoring reports assessing poorly conducted elections on the one hand, and ten reports that do the same thing for elections with reportedly highest standards on the other (see Table A1). Following section summarizes the main analytical steps (Part 1 & Part 2) and presents them as an algorithmic description of the whole approach.

Part 1: Data preparation

1. Each document is manually cleaned off of its title page, content, appendix, footers, and headers.
2. All documents are analyzed for presence of named entities (e.g. locations, organizations, people). These entities are properly marked so they can be later removed as a potential source of bias introduced to the scaling algorithm.
3. All documents are lemmatized (a process of grouping together the inflected forms of a word so they can be analyzed as a single item).
4. Documents are cleaned off of stop-words, infrequent words (occurring less than 5 times across corpus), numbers, punctuation, whitespaces, and named entities.
5. All upper case characters are transformed to lower case.
6. Cleaned and preprocessed corpus is used in Part2: Data analysis.

Part 2: Data analysis

1. After examining all collected reports using close reading, twenty of them are selected for sampling (see Table A1).
2. The reports are divided into two groups with indexes denoting their affiliation to low quality (lq) $\{lq_1, lq_2, \ldots, lq_{10}\}$ and high quality (hq) $\{hq_1, hq_2, \ldots, hq_{10}\}$ elections.
3. Rather than grouping the reports, we opt for individual pairs of documents $(lq_n - hq_n)$ and sample all possible pairs (n = 100) given by the selected subset; *pairs* = $\{lq_1 - hq_1, lq_2 - hq_1, lq_3 - hq_1, \ldots, lq_{10} - hq_8, lq_{10} - hq_9, lq_{10} - hq_{10}\}$.
4. Having 100 unique pairs of reference documents, we train 100 wordscores models (using default settings in *quanteda* package) and use the pairs of sampled documents as seeds for producing scaling scores for all the remaining virgin documents across 100 models. Additional benefit of this approach is that we do not have to discard the reference documents or artificially rescaled them as they are rotated using bootstrapping. Their scores are then stabilized through iterations in which they are not part of the reference pairs (using 90 instead of 100 scores).

   Formally:

   > **for each** *pair* in *pairs* {
   > > **train** a wordscores model with lq_{pair_n} and hq_{pair_n} as reference documents
   > > **predict** scores of the virgin documents
   > > **save** the scores
   > }

5. After training 100 models with 100 unique pairs of reference documents, the result is a set of 100 wordscores scales which map the position of all scaled documents across different pairs of reference documents.
6. By averaging the scores across all trained models, we get a stabilized position of a document on a latent scale defined by multiple reference documents.

Notes

1 Benoit et al., "Quanteda: An R Package for the Quantitative Analysis of Textual Data."
2 Benoit and Laver, "Estimating Irish Party Policy Positions Using Computer Wordscoring: The 2002 Election – a Research Note"; Lowe, "Understanding Wordscores"; Bruinsma and Gemenis, "Validating Wordscores: The Promises and Pitfalls of Computational Text Scaling."

Turnover election

Coding of turnover election [0-1] draws on conceptualisation proposed by Judith Kelley[1] with few modifications:

1) Under a parliamentary system, the variable is coded as 0 if the incumbent political party (or a coalition partner) remains in the government after the election, or if a president from the ruling party (or supported by the ruling party) wins the election. In parliamentary systems with collective presidency (namely Bosnia and Herzegovina), the variable is coded as 1 when two presidential candidates are elected from parties which were not part of the ruling coalition before election;

2) Under a presidential system, the variable is coded as 0 if there is a presidential election only and the incumbent party retains the office of the president. If there is a legislative election only, the variable is coded as 0 if the presidential party wins the election. If the parliament comprises two chambers (US), the variable is coded as 1 if both chambers are won by the opposition party, i.e. party other than the presidential party. If there is a general election, the variable is coded as 0 if the incumbent leading party in the legislature maintains its leading role and the incumbent party of the president retains its incumbency. However, because it is a presidential system, the variable is also coded as 0 if there is a general election and the incumbent party of the president retains its incumbency, even if the legislative lead party loses its plurality;

3) In single-party or no-party states, the variable is generally coded as 0. This is also true if a candidate is installed as acting president prior to the election and then wins the election, or if a chosen presidential successor is elected to replace the president.

Transitional election

Transitional election [0-1] is coded in line with conceptualizations provided by International IDEA[2]. The framework identifies three categories of transitional elections:

1) Shift from authoritarianism or semi-authoritarianism to electoral democracy.
   - These countries will likely have no electoral traditions or will hold elections that emulate a democratic process in order to keep a long-standing semi/authoritarian regime in power. The shift towards electoral democracy is precipitated by internal factors (such as popular uprisings) or external ones (such as economic sanctions);
2) Transition from deep political crises to political stability.
   - Some countries experience deep political crises resulting from unconstitutional changes of government, disputed election results, authoritarian tendencies of newly elected governments, or other major political divides. Protracted crises may cause the meltdown of institutions responsible for their resolution (parliament and government), and delaying the conduct of extraordinary elections beyond deadlines set in the constitution and laws;
3) Transition from war to peace.
   - Peace processes often entail considerations on when to organize elections with the aim of restoring or establishing democratic governance. Such provisions are often stipulated in peace agreements.

Mission size

Mission size [1-5] is coded based on the number of election observers deployed to a country relative to the country's size. As for the country size, countries are divided into three categories (small, medium, large) depending on the size of their population, reflecting the population scope in the OSCE states: small (0-4 million); medium (4-40 million); large (40 million and more). A few exceptions exist where due to the vast geographical size combined with a low population density the country is assigned to a higher category of country size despite lower population (Afghanistan, Canada, Kazakhstan, Mongolia, Uzbekistan). Five different categories of mission size are then coded relative to the country size (smallest [1], small [2], medium [3], large [4], largest [5]). Number of observers is a cumulative number of all long-term and short-term observers deployed to the country, reflecting the real numbers of observers deployed. Based on these criteria, the mission size is coded as follows. For small countries: smallest (core team only); small (2-132 observers); medium (133-264); large (265-396); largest (397-528). For medium countries: smallest (core team only); small (2-165); medium (166-330); large (331-495); largest (496-658). For large countries: smallest (core team only); small (2-330); medium (331-660); large (661-990); largest (991-1318). In practical terms, missions of limited format where only core team of experts is deployed are always coded as 'smallest', i.e. election assessment missions, election expert teams, and election support teams.

Notes

1   Kelley, Monitoring Democracy: When International Election Observation Works, and Why It Often Fails, 168–78.
2   Alihodžić et al., "Timing Seq. Transitional Elections Int. IDEA Policy Pap. No. 18," 11–12.

Example #1: Lenient rhetoric towards Russian Federation

An often-cited example of compromised language can be found in the OSCE/ODIHR election observation report on the 2000 presidential election in Russia. While various sources brought up attention to voting irregularities and unequal access to media[1], pervasive and decisive fraud, and allegations about 1.3 million names illegally added to voter lists[2], the OSCE/ODIHR preliminary report limited itself to a statement that "*[t]he irregularities noted in the polling and the vote count did not appear to have an impact on the outcome of the election*".[3] Although the following final report backed off from such lenient positions, a restrained wording remained in the text, pointing out that "*[i]n general, and in spite of episodic events that sometimes tested the system's capacity to uphold principles of fairness and a level playing field, the presidential election was conducted under a constitutional and legislative framework that is consistent with internationally recognized democratic standards*".[4] The report concluded that "*[t]his election also demonstrated Russia's continuing commitment to strengthen its democratic electoral institutions, which appear to have the public's confidence and acceptance*" (ibid.).

Example #2: Lenient rhetoric towards Azerbaijan

The OSCE/ODIHR was also criticised for lenient assessment of the 2008 presidential elections in Azerbaijan. According to Freedom House, the elections were "clearly uncompetitive".[5] Domestic opposition media reported that the poll was a "show" marking the "most antidemocratic" election in the country's history.[6] Yet, the OSCE/ODIHR monitoring mission in its preliminary statement declared that "*the elections marked considerable progress toward meeting OSCE and Council of Europe commitments and other international standards*".[7] This was criticized by both the domestic as well as the international actors. EU council pointed out that "*the elections still do not satisfy international standards of democracy, particularly as regards the organization of public debate, the conduct of polling, and the counting of votes*".[8] After a wave of criticism, OSCE/ODIHR backed off from its previous praising positions stating that "*[w]hile the presidential election marked considerable progress towards meeting OSCE commitments and other international standards, in particular with regard to some technical aspects of election administration, the election process failed to meet some OSCE commitments*".[9] Although the report provided a more critical assessment, it was still far from the harsh criticism voiced by other international and domestic actors.

Example #3: Lenient rhetoric towards Georgia

The last example of criticism towards OSCE/ODIHR monitoring mission covers the 2020 parliamentary elections in Georgia when both the preliminary statement and the final report stated that *"[t]he 31 October parliamentary elections were competitive and, overall, fundamental freedoms were respected".*[10] Although harsher wording followed in sentences after this opening line, the assessment was widely considered as compromising the overall tone of the report. Freedom House noted that *"[the r]uling party members capitalized on the report's positive opening line and interpreted it publicly as representing international recognition of the elections".[11]*. At the same time, the opposition parties proclaimed elections as rigged and refused to accept the outcome.[12] The conduct of elections received extensive domestic and some international criticism.[13] According to local election-monitoring groups the elections were *"the least democratic and free among elections"* held under the incumbent party.[14] The Freedom House

concluded that the elections marked the end of the country's democratic reforms and led to democratic backsliding strongly resembling the situation a decade ago.[15]

Notes

1   Freedom House, "Freedom in the Wordl 2000-2001," 444–50.
2   Mendelson, "Democracy Assistance and Political Transition in Russia: Between Success and Failure," 102–3.
3   OSCE/ODIHR, "Russia: Presidential Election, 26 March 2000 - Preliminary Statement," 3.
4   OSCE/ODIHR, "Russia: Presidential Election, 26 March 2000 - Final Report," 3.
5   Freedom House, "Nations in Transit 2009: Democracy's Dark Year," 7–8.
6   Congressional Research Service, "Azerbaijan's October 2008 Presidential Election: Outcome and Implications," 4.
7   OSCE/ODIHR, "Statement of Preliminary Findings and Conclusions on the 15 October 2008 Presidential Election in Azerbaijan," 1.
8   Congressional Research Service, "Azerbaijan's October 2008 Presidential Election: Outcome and Implications," 4.
9   OSCE/ODIHR, "Final Report on the 15 October 2008 Presidential Election in Azerbaijan," 1.
10 OSCE/ODIHR, "Georgia Parliamentary Elections, 31 October 2020 - Preliminary Report," 1; OSCE/ODIHR, "Georgia Parliamentary Elections, 31 October 2020 - Final Report," 2.
11 Freedom House, "Georgia: Nations in Transit 2021 Country Report."
12 JAM News, "Opposition Parties in Georgia Sign Joint Statement Renouncing Their Seats in Parliament."
13 Freedom House, "Georgia: Nations in Transit 2021 Country Report."
14 Georgian Young Lawyers' Association, "NGOs Assessment of the 2020 Parliamentary Elections - Georgian Young Lawyers' Association."
15 Freedom House, "Nations in Transit," 4.

References (Appendix)

Alihodžić, Sead, Nicholas Matatu, Oliver Joseph, and Katrin Lewis. "Timing and Sequencing of Transitional Elections: International IDEA Policy Paper No. 18." *Timing and Sequencing of Transitional Elections: International IDEA Policy Paper No. 18*. International IDEA Policy Paper. International Institute for Democracy and Electoral Assistance, 2019. https://doi.org/10.31752/IDEA.2019.13.

Benoit, Kenneth, and Michael Laver. "Estimating Irish Party Policy Positions Using Computer Wordscoring: The 2002 Election – a Research Note." *Irish Political Studies* 18, no. 1 (2003): 97–107. https://doi.org/10.1080/07907180312331293249.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3, no. 30 (October 6, 2018): 774. https://doi.org/10.21105/joss.00774.

Bruinsma, Bastiaan, and Kostas Gemenis. "Validating Wordscores: The Promises and Pitfalls of Computational Text Scaling." *Communication Methods and Measures* 13, no. 3 (2019): 212–27. https://doi.org/10.1080/19312458.2019.1594741.

Congressional Research Service. "Azerbaijan's October 2008 Presidential Election: Outcome and Implications," 2008.

Freedom House. "Freedom in the Wordl 2000-2001," 2001.

———. "Georgia: Nations in Transit 2021 Country Report," 2021. https://freedomhouse.org/country/georgia/nations-transit/2021.

———. "Nations in Transit." *2021*, 2021.

———. "Nations in Transit 2009: Democracy's Dark Year," 2009.

Georgian Young Lawyers' Association. "NGOs Assessment of the 2020 Parliamentary Elections - Georgian Young Lawyers' Association," 2020. https://gyla.ge/en/post/arasamtavrobo-organizaciebi-2020-is-tslis-saparlamento-archevnebs-afaseben#sthash.FVRM6nUn.ZnKm7wTu.dpbs.

JAM News. "Opposition Parties in Georgia Sign Joint Statement Renouncing Their Seats in Parliament," 2020. https://jam-news.net/georgia-document-nika-melia-opposition-parliament-boycott-national-movement-lelo/.

Kelley, Judith G. *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*. Princeton: Princeton University Press, 2012.

Lowe, Will. "Understanding Wordscores." *Political Analysis* 16, no. 4 (2008): 356–71. https://doi.org/10.1093/pan/mpn004.

Mendelson, Sarah E. "Democracy Assistance and Political Transition in Russia: Between Success and Failure." *International Security* 25, no. 4 (2001): 68–106.

OSCE/ODIHR. "Final Report on the 15 October 2008 Presidential Election in Azerbaijan," 2008. https://www.osce.org/files/f/documents/5/3/35625.pdf.

———. "Georgia Parliamentary Elections, 31 October 2020 - Final Report," 2020. https://www.osce.org/files/f/documents/1/4/480500.pdf.

———. "Georgia Parliamentary Elections, 31 October 2020 - Preliminary Report," 2020. https://www.osce.org/files/f/documents/a/d/469005.pdf.

———. "Russia: Presidential Election, 26 March 2000 - Final Report," 2000. https://www.osce.org/files/f/documents/3/9/16275.pdf.

———. "Russia: Presidential Election, 26 March 2000 - Preliminary Statement," 2000.

———. "Statement of Preliminary Findings and Conclusions on the 15 October 2008

Presidential Election in Azerbaijan," 2008.
https://www.osce.org/files/f/documents/7/a/34525.pdf.