Measuring Likely Voters: A Model for Predicting Electoral Participation in Romanian Presidential Elections
Sultanescu, Dan; Buti, Daniel; Sultanescu, Dana

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Article

Measuring Likely Voters. A Model for Predicting Electoral Participation in Romanian Presidential Elections

Dan SULTĂNESCU

Center for Civic Participation and Democracy (CPD),
National University of Political Studies and Public Administration,
Research Director, Bucharest, Romania;
dan.sultanescu@snspa.ro

Daniel BUTI

National University of Political Studies and Public Administration,
lecturer, Bucharest, Romania;
daniel.buti@snspa.ro

Dana SULTĂNESCU

Center for Civic Participation and Democracy (CPD),
National University of Political Studies and Public Administration,
expert, Bucharest, Romania;
dana.sultanescu@drd.snspa.ro

Abstract: *Predicting voter turnout by estimating the proportion of likely voters as accurately as possible is a challenging task and it is debated both in literature and among experts actively involved in the field of sociological research and polling organizations. While there are several variables that influence turnout, and also several methods and models that have been put into practice over time, in reality there is no magic formula for correctly measuring, before an election, the percentage of individuals that are most likely to vote. Each national context comes with particular variables, as well as historical factors and demographic dynamics that need to be taken into account. We propose a model built on pre-election survey data and on past official election statistics, which has been tested and proven viable in the context of the 2019 Romanian presidential election.*

Keywords: *electoral participation; likely voters; predicting turnout; Romania*

1. Introduction

Estimating voter turnout is an issue of high interest for both researchers and political actors. Electoral participation is an expression of the democratic commitment, of people's interest in elections, of mobilization or dissatisfaction / disappointment towards the political system and its actors (Kaase, 2007), influencing political calculations and electoral strategies. Pre-election measurement and especially predicting turnout as accurately as possible has always been a challenge for pollsters and political scientists, given not only the issue of survey respondents' over-estimation of voting intentions, but also the multiple variables that can be taken into account for building a predictive model.

Literature, but also research institutes offer different formulas or good practices, but their applicability is not a general one, at least not without nuances and adjustments to specific contexts and situations. Therefore, the model this paper proposes is based on experience and knowledge accumulated over time in this field, enriched with specific characteristics of the Romanian elections, especially presidential elections. This model is built on pre-election survey data and on official election statistics, being tested in the context of the 2019 Romanian presidential election[1].

Model design involves a mix of variables and a step-by-step development, by weighting and recalibrating the data. It involves two categories of variables: poll questions – intention to vote, voting in the past, interest in elections (interest in potential elections in the near future), and official election data (turnout in different types of elections from the past).

From a methodological point of view, there are several stages involved in the construction of the model for predicting electoral participation. The first one is to define the main variable – in our case, electoral participation (intention to vote). Next, it is essential to identify useful and relevant types of *reference behavior*. These are variables that can generally be associated with electoral participation and voting intention, such as past voting behavior, previous behavior involving different types of political participation (signing petitions, displaying a campaign badge, donating money to a political organization or group, participating in political activities, taking part in protests etc.), attitudes and orientations (values, feelings, opinions, etc.) that may predict a stronger or weaker willingness to participate in voting and public life in general (Inglehart, 1997; Norris, 2002; Norris, 2007; Oser, 2017; Inglehart, 2018; Norris and Inglehart, 2019), relevant statistics that may be used to understand participation, such as census data, official government statistics or other data. The third stage involves performing data analysis and building the actual model. To this end, simple bivariate analysis can be used, which can create linear models between two variables, or more complex, multivariate analyzes can be run, such as regressions or cluster analysis, as well as other types of segmentation of the populations based on complex variation of the variables. Data analysis allows the identification of the most relevant variables in the data set that can be considered when constructing the prediction model. Defining the model involves taking into account certain variables and excluding others, namely determining those independent variables that best capture the behavior to be predicted. The final step is to refine the model by including in the analysis certain sampling limitations, but also official electoral statistics.

2. Measuring the proportion of likely voters. A brief literature review

Over-reporting of voting in a survey context is a well-known issue among social science researchers. This phenomenon can have important consequences for research that seeks to explain and / or estimate turnout. It can distort social reality, influencing research hypotheses (Bernstein et al., 2001) and directing researchers towards low-relevance paths. There are several significant factors that contribute to the gap between declared intention to vote and actual turnout, as a result of respondents in post-electoral surveys who report having voted despite the fact that they did not participate, or citizens who report before the election that they intend to vote, but they do not do so on election day. The respondents' memory plays a role in this process, as does social pressure / desirability, or in other words, social desirability response

bias (Burden, 2000; Karp and Brockington, 2005; Duff et al., 2007; Gerber and Rogers, 2009; Holbrook and Krosnick, 2010). When people feel that they will be judged or have something to lose due to their opinion, they tend to give untrue but socially acceptable responses (Bernstein et al., 2001).

Studies of this phenomenon have sought to reveal its correlations with various socio-demographic and attitudinal variables of respondents, in order to be able to "identify" and "isolate" those who tend to over-report voting (Bernstein et al., 2001; Silver et al., 1986; Belli et al., 2001; Deufel and Kedar, 2010; Stout and Martin, 2016). Others focused on the type of interview, the improvement of data collection procedures or the development of the survey tool, more specifically the design of survey questions (Duff et al., 2007; Gerber and Rogers, 2009; Holbrook and Krosnick, 2010; Katz and Katz, 2010; Hanmer et al., 2014; Comşa and Gheorghiţă, 2016), to reduce social desirability bias and to increase the accuracy of the responses.

The official results of electoral participation are usually compared with sociological data collected (by opinion poll) either in the run-up to the election – data that reflect a particular voting intention – or after the election – data that highlight the inclination to, and dimension of the conformist response. Since estimated voting intention does not always materialize, for polling organizations, social sciences experts, and researchers, but also for political actors, the first type of comparison often poses a real challenge: developing better methods for estimating turnout as accurately as possible, in other words a precise quantitative determination of the percentage of *likely voters* – people who have a strong intention to vote, who are the most likely to put into action their intention expressed in an opinion poll.

There is no magic formula for measuring, before the election, those voters with the highest propensity to participate. Research institutes use a number of indicators, but their combination differs. As such, there is no right or wrong methodology for determining the most likely turnout (AAPOR, n.d.). For example, institutes such as Gallup or Pew Research use the Perry-Gallup index. These questions focus on intention to vote, past voting behavior, knowledge about the voting process and interest in the election campaign. Other models also add socio-demographic data about respondents (Perry-Gallup and Demographics – PGaD) such as age, education, income, gender, strength of partisanship, etc. (Rentsch et al., 2019). Pre-election information is aggregated and processed, generally using deterministic methods to divide survey respondents into voters and non-voters, or probabilistic methods, to calculate the probability of each respondent going to vote (Keeter and Igielnik, 2016; Rentsch et al., 2019).

Estimating electoral participation is a process that involves managing several variables. The sociological data produced by questionnaires provide relevant information, but the actual participation (materialization of the intention to vote, its transformation into action) can be influenced by several factors recorded by the rich literature on voting behavior. Among these are institutional arrangements, such as compulsory voting or the type of electoral system; the importance / the relevance of the elections; socio-economic factors such as population size, income inequality or corruption; circumstantial factors, such as the competitiveness of the party system and the balance of power between electoral actors (Geys, 2006; Blais, 2006; Cancela and Geys, 2016; Stockemer, 2017).

3. Research design. Predicting the voting behavior of the population – electoral participation, based on survey data, in Romania

The construction of the model for predicting electoral participation is based on the data of an opinion poll conducted by CPD – the Center for Civic Participation and Democracy (a research center of the National University of Political Studies and Public Administration – SNSPA) in 2019, five months before the presidential election. The survey data were collected from June 1 to 22 (N=977, ±3%). The analysis also uses statistics from previous elections, from the national electoral bureau in Romania.

3.1. The 1ˢᵗ stage of the process. Defining the variable

Since our main variable is electoral participation, we first look at the percentage of people who declare their intention to vote in the future presidential elections, as a response to a direct question in our survey, and discover that 83% of the respondents indicate intention to participate when asked straightforwardly (Fig. 1). The percentage expresses the answer to a standard, general question, frequently biased by social desirability, and our objective is further refining this number to better reflect reality.
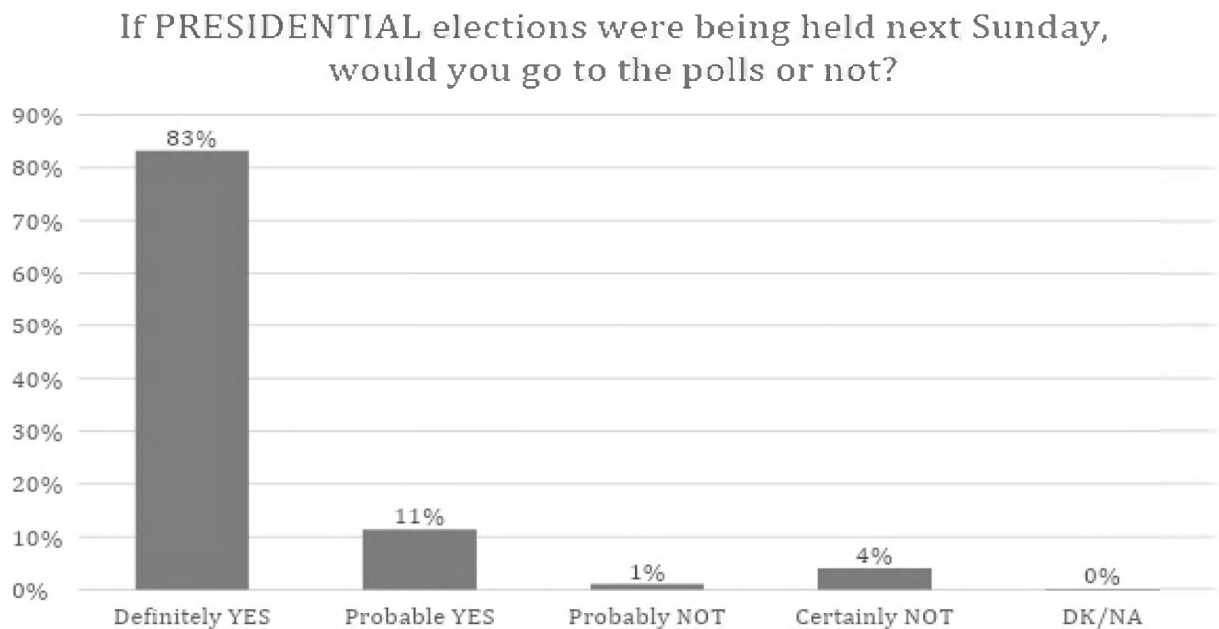
Figure 1. Distribution of responses to the direct question estimating electoral participation

3.2. The 2ⁿᵈ stage of the process. Identifying reference behavior

In this process of refining and fine-tuning the model, we need to identify similar behaviors that can be compared and interpreted to extract useful data in the process of predicting electoral participation. Our initial choice is to evaluate past behavior involving the main variable – i.e., past

voting episodes, operating on the premise that "voting in one election substantially increases the likelihood of voting in the future" (Gerber et al., 2003, p. 540), that "most people are habitual voters" (Fowler, 2006, p. 3) or in other words, voting in the past increases the probability of voting in the future. Assessment of past electoral behavior can use various sources of data: on the one hand, appealing to the memory of the survey respondents, and, on the other hand, appealing to official data on actual participation in various types of past elections.

Survey data show that 81% of respondents participated in the presidential election in the past (Fig. 2). This percentage is a new benchmark, an additional piece of information that contributes to the construction of the model. But memory is selective and sometimes inaccurate. In addition, due to social pressure, people tend to sometimes give responses that are expected from them (more people say they voted compared to actual participation figures).
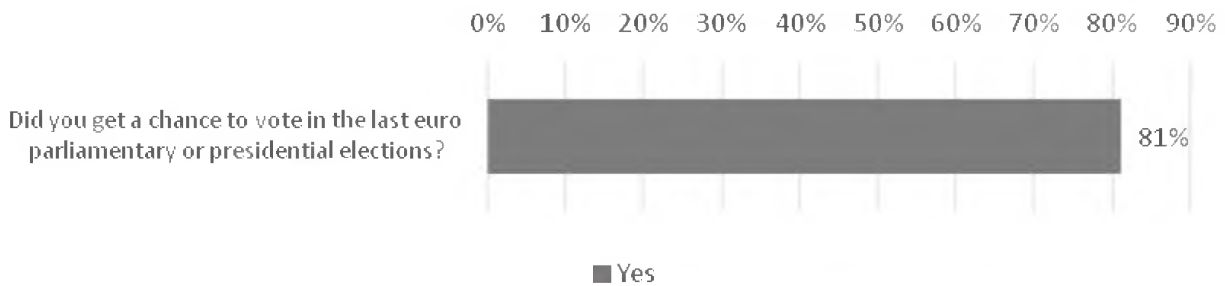

Figure 2. Subsample of population responding „Yes" to the question
regarding voting in recent past elections

Comparing this percentage with historical data on past turnout at different elections shows that there is a considerable difference between official numbers and what people say, with actual turnout being lower (Table 1). Indeed, no election moment resembles another, and the perceived stakes of the electoral process, as well as the importance given to each type of election (local, regional, national, European) influence turnout. However, the differences are significant, and they require a refinement of the model.

Table 1. Electoral participation in Romania. Official turnout for several rounds of elections

Type of election	Date	Electoral participation rate
Presidential elections	November 2014	53.17%
Local elections	June 2016	48.17%
Parliamentary elections	December 2016	39.44%
European Parliament elections	May 2019	49.65%

Thus far, our sample reveals 83% declared intention to participate, an indicator that we cannot assume to be reliable, since only 81% of the respondents say they voted in the previous presidential election, and we also know from official data that turnout actually ranged from 39% to 53%.

In order to get a more accurate estimation of the percentage of people who actually intend to vote, we need to complete the model by introducing new relevant variables. Specifically, we build an alternate predictor of voting, using a combination of two variables: intention of participating, our main variable (with a scale of four answers: *definitely YES, probably YES, probably NOT, definitely NOT*) and interest in elections (also using a scale of four possible answers: *very interested, somewhat interested, not very interested, not at all interested*).

If PRESIDENTIAL elections were being held next Sunday
(for the election of the President of Romania), how
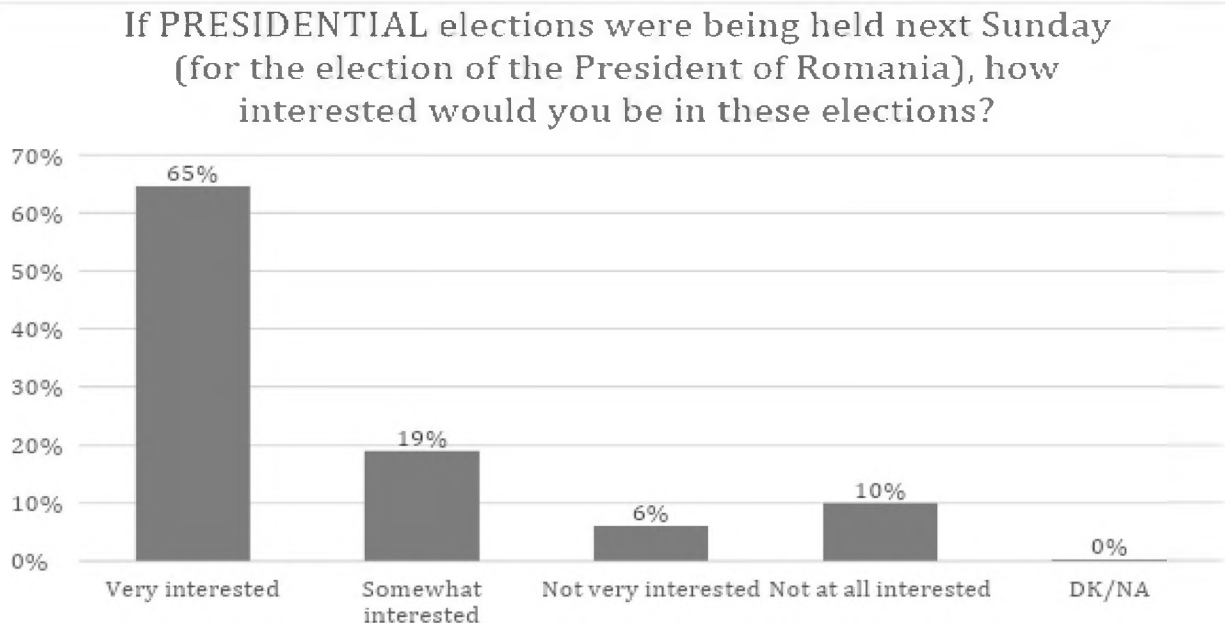interested would you be in these elections?

Figure 3. Distribution of responses to the question estimating interest in elections

At this point, our sociological data show an 83% stated intention to vote (Fig. 1) and an interest in hypothetical (presidential) elections of 65% (Fig. 3). The combination of these data can provide a clearer picture of voter turnout.

3.3. The 3rd stage of the process. Data analysis and model-building

By using a simple crosstabulation between these two variables, we can obtain a useful segmentation of the sample. We find a hierarchy of several categories of respondents, based on the intensity of their willingness to participate (Table 2): the first category of very mobilized voters is composed of people that say they will definitely vote, and (at the same time) they are very interested in elections; subsequently, we have various levels of mobilization – people with slight interest in election, or only mild intention to vote; finally, there are individuals that have either no intention to vote, or no interest in the elections in general.

Table 2. Simple crosstabulation highlighting the relationship between the two questions – interest in elections and intention to vote

		If PRESIDENTIAL elections were being held next Sunday, would you go to the polls or not?					Total
		Definitely YES	Probably YES	Probably NOT	Certainly NOT	DK/NA	
If PRESIDENTIAL elections were being held next Sunday (for the election of the President of Romania), how interested would you be in these elections?	Very interested	62%	2%	0%	0%		**65%**
	Somewhat interested	17%	2%				**19%**
	Not very interested	2%	3%	0%	1%	0%	**6%**
	Not at all interested	3%	3%	1%	3%		**10%**
	DK/NA	0%	0%				**0%**
Total		**83%**	**11%**	**1%**	**4%**	**0%**	100%

So far, we can assume that the people who are extremely interested in elections and say they "certainly" intend to vote have the highest probability to vote. Therefore, from a subsample of 81% of people that said they voted in the past, we manage to narrow the search for "likely voters" to a lower sample of 62%, which is a more plausible participation rate. Nonetheless,

this figure is still much higher compared to the real-life general participation behaviors (which are, usually, below 50%).

To refine the predictive model, the data must be further adjusted, using information about similar past behavior. To this end, we take into account answers to a question regarding past participation to elections (with a simple response option – *Yes* or *No*). Since the survey was conducted only a few weeks after the elections for the European Parliament (in May 2019), the results are relevant.

We create a new crosstabulation report to explore the relationship between the two variables, namely the response of the sample of 62% (people who seem to be very mobilized) in terms of their voting behavior in the most recent past elections. The result shows a subsample of 55% of "likely voters" (Table 3), people who, at the same time, are very mobilized to vote in the presidential elections and have voted in recent past elections. We can look at this number with increased confidence that it resembles more the percentage of people that will definitely go to vote on the day.

Table 3. Simple crosstabulation highlighting the relationship between the two variables – mobilization for presidential elections and voting behavior during previous elections

		Did you get a chance to vote in the last euro parliamentary or presidential elections?		
		No	Yes	Total
Mobilization for presidential election	Very mobilized	7%	55%	**62%**
	Slightly mobilized	6%	16%	**21%**
	Not mobilized	6%	10%	**17%**
Total			**19%**	**81%**	**100%**

3.4. The 4th stage of the process. Refining the data. Taking Diaspora into account to adjust the general population level

This step involves an adjustment of the survey data to the reality of the actual population, more precisely to the official statistics about the electorate. In any national sample, there is no technical possibility to include people living or working abroad (the so-called Diaspora). This generates an error from the start, and it is necessary to adjust the results to the current adult population since there are estimates (NIS, 2019; Toma, 2020; OECD, 2019) which suggest that people currently living in Romania only represent 85% of the adult population (thus, our survey sample is not representative for all 18.3 million Romanian adults, but for 85%*18.3=15.6 million adults actually living in the country, while around 2.7 million people live and work abroad).

Consequently, we need to adjust our prediction for two different segments of the population. First, for the people that vote in Romania, and secondly, for the people that vote abroad.

Taking this into consideration, we can presume that, if the 55% turnout is representative for the *in-country* population, then, when applied to the *whole voting population*, this percentage will decrease to 47%, which is our prediction for the turnout of the population living in Romania.

For the Diaspora turnout, we first need a historical perspective. Analyzing past voting behaviors by studying official turnout data, we find that in 2014-2019 the number of Romanians voting abroad varied between 106.038 for the 2016 parliamentary elections – BEC (n.d.-a), 161.262 at the 2014 presidential elections – BEC (n.d.-b), and 384.943 people for the 2019 European Parliament elections – BEC (n.d.-c), which represents a 1-2% bonus to the overall turnout. By adding these percentages, we can estimate a turnout in the presidential election between 48% and 49%. This is the last stage of the process and the last piece of the model of predicting electoral participation, which introduces one last variable and refines the entire process.

4. Results. How well did the predicting model perform compared to actual data?

Our estimation of turnout was based on data measured in June 2019, a few months before the actual election day (10 November 2019), and just after another round of elections (for the European Parliament – 26 May 2019). The real turnout in Presidential elections, in November 2019, was 47.66% in Romania and 3,67% abroad (among the Diaspora).

This means that the total participation rate was 51.33%. comparatively (Table 4), our estimate was 48-49%.

Table 4. Actual electoral participation vs. estimated turnout based on our predictive model (2019 Romanian presidential elections)

Official turnout	47.66%
Total turnout (RO + Diaspora)	51.33%
Estimation	47%
Difference 1 (Official turnout – Estimation)	- 0.66%
Difference 2 (Total turnout – Estimation)	2.83%

The margin of error for the model we used proved to be very small for the "domestic" population (only 0,66%), and slightly larger for the turnout including the diaspora (around 2,8%). Taking into account that our data was collected only from people living on the Romanian territory, this method of estimating the voting participation has proven very effective.

5. Discussion

This heuristic method has been tested repeatedly in different election periods (2014, 2016, 2020), with similar results. Our experience shows that, using this combination of variables, we can predict not only the level of turnout, but also the results of the elections (with scores for every party/candidate with a similar margin of error).

Voting can be influenced by many factors, some of which are subjective (related to each individual), others specific to the political and social system. Therefore, the actual turnout varies from one electoral moment to another, and its preliminary estimation cannot be a simplistic process. Predicting voting turnout as accurately as possible must be based on the experience of the researcher or pollster, on a series of good practices in the field, but also on building an appropriate model of secondary analysis involving the choice of the most relevant variables and their careful calibration.

The model presented here does not presume to be infallible, but it has the ability to increase the accuracy of the pre-election polls, with direct implications for public confidence in this instrument, in the basic foundation of the electoral system, and the effectiveness of electoral campaigns. Ultimately, the importance and relevance of any predictive model is given by its confrontation with reality, in this case its comparison with the official turnout.

Notes

[1] This model is the result of a heuristic process, and the practical activity, on an empirical level, of sociologists Vlad Achimescu, Leonard Sultănescu and Mirel Palada who contributed to its development over time, for which the authors are grateful.

Acknowledgements

Conflicts of interest

The authors declare no conflict of interest.

References

American Association for Public Opinion Research (AAPOR) (n.d.) *Likely Voters*. Available at: https://www.aapor.org/Education-Resources/Election-Polling-Resources/Likely-Voters.aspx [Accessed on 18 May 2022].

BEC (n.d.-a) *Alegeri parlamentare decembrie 2016*. Available at: https://prezenta.bec.ro/parlamentare2016/ [Accessed on 04 July 2022].

BEC (n.d.-b) *Situația prezenței la urne la*

alegerile pentru Președintele României – 2 noiembrie 2014. Available at: http://bec2014.roaep.ro/wp-content/uploads/2014/11/SIAP2014_PAR_Raport-Situatie-Prezenta-la-urne.pdf [Accessed on 04 July 2022].

BEC (n.d.-c) *Alegeri Europarlamentare 26 mai 2019*. Available at: https://prezenta.bec.ro/europarlamentare 26052019/abroad-pv-final [Accessed on 04.07.2022].

Belli, R.F., Traugott, M.W. and Beckmann, M.N. (2001) What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies. *Journal of Official Statistics*, 17(4), pp. 479-98.

Bernstein, R., Chadha, A. and Montjoy, R. (2001) Overreporting Voting: Why it Happens and Why it Matters. *Public Opinion Quarterly*, 65(1), pp. 22-44.

Blais, A. (2006) What Affects Voter Turnout? *Annual Review of Political Science*, 9(1), pp. 111-125.

Burden, B.C. (2000) Voter turnout and the national election studies. *Political Analysis* 8(4), pp. 389-398.

Cancela, J. and Geys, B. (2016) Explaining Voter Turnout: A Meta-Analysis of National and Subnational Elections. *Electoral Studies*, 42, pp. 264-275.

Comșa, M. and Gheorghiță, A. (2016) Many', 'Half' or 'One out of Two'? Assessing the Effectiveness of Counter-Biasing Techniques in Reducing Self-Reported Turnout. *Revista de cercetare și intervenție social*, 55, pp. 80-97.

Deufel, B.J. and Kedar, O. (2010) Race and Turnout in U.S. Elections: Exposing Hidden Effects. *Public Opinion Quarterly*, 74(2), pp. 286-318.

Duff, B., Hanmer, M.J., Park, W-H. and White, I.K. (2007) Good Excuses: Understanding Who Votes with an Improved Turnout Question. *Public Opinion Quarterly*, 71(1), pp. 67-90.

Fowler, J.H. (2006) Habitual Voting and Behavioral Turnout. *Journal of Politics*, 68(2), pp. 335-344.

Geys, B. (2006) Explaining Voter Turnout: A Review of Aggregate-level Research. *Electoral Studies*, 25(4), pp. 637-663.

Gerber, A.S., Green, D.P. and Shachar, R. (2003) Voting May Be Habit Forming: Evidence from a Randomized Field Experiment. *American Journal of Political Science*, 47(3), pp. 540-550.

Gerber, A.S. and Rogers, T. (2009) Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You. *The Journal of Politics*, 71(1), pp. 178-191.

Hanmer, M.J., Banks, A.J. and White, I.K. (2014) Experiments to Reduce the Over-reporting of Voting: A Pipeline to the Truth. *Political Analysis*, 22(1), pp. 130-141.

Holbrook, A.L. and Krosnick, J.A. (2010) Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique. *The Public Opinion Quarterly*, 74(1), pp. 37-67.

Inglehart, R. (1997) *Modernization and postmodernization: Culture, economic and political change in 43 societies*. Princeton, NJ: Princeton University Press.

Inglehart, R. (2018) *Cultural evolution. People's Motivations are Changing, and Reshaping the World*. Cambridge: Cambridge University Press.

Kaase, M. (2007) Perspectives on Political Participation. In: R.J. Dalton and H-D. Klingemann, eds. *The Oxford Handbook of Political Behavior*. Oxford: Oxford University Press, pp. 783-796.

Karp, J.A. and Brockington, D. (2005) Social desirability and response validity: a comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics*, 67(3), pp. 825-840.

Katz, J.N. and Katz, G. (2010) Correcting for Survey Misreports Using Auxiliary Information with an Application to Estimating Turnout. *American Journal of Political Science*, 54(3), pp. 815-835.

Keeter, S. and Igielnik, R. (2016) Can Likely Voter Models Be Improved? Evidence from the 2014 U.S. House elections. *Pew Research Center*, January 7. Available at: https://www.pewresearch.org/methods/2016/01/07/can-likely-voter-models-be-improved/ (Accessed on 23.05.2022).

National Institute of Statistics (NIS). (2019) Available at: *Studiu exploratoriu privind stocurile de migra?ie* (*Exploratory study on migration stocks*), March 1. Available at: https://insse.ro/cms/en/node/5642 [Accessed on 04 July 2022].

Norris, P. (2002) *Democratic Phoenix: Reinventing Political Activism*. Cambridge: Cambridge University Press.

Norris, P. (2007) Political Activism: New Challenges, New Opportunities. In: C. Boix and S.C. Stokes, eds. *The Oxford Handbook of Comparative Politics*. Oxford: Oxford University Press, pp. 628-652).

Norris, P. and Inglehart, R. (2019) *Cultural backlash: Trump, Brexit, and the rise of authoritarian-populism*. New York, NY: Cambridge University Press.

OECD (2019) *Talent Abroad: A Review of Romanian Emigrants*, July 16. Available at: https://www.oecd.org/publications/talent-abroad-a-review-of-romanian-emigrants-bac53150-en.htm [Accessed on 04 July 2022].

Oser, J. (2017) Assessing How Participators Combine Acts in Their "Political Tool Kits": A Person-Centered Measurement Approach for Analyzing Citizen Participation. *Social Indicators Research*, 133(1), pp. 235-258.

Rentsch, A., Schaffner, B.F. and Gross, J.H. (2019) The Elusive Likely Voter: Improving Electoral Predictions with More Informed Vote-Propensity Models. *Public Opinion Quarterly*, 83(4), pp. 782-804.

Silver, B.D., Anderson, B.A. and Abramson, P.R. (1986) Who overreports voting? *American Political Science Review*, 80(2), pp. 613-624.

Stockemer, D. (2017) What Affects Voter Turnout? A Review Article/Meta-Analysis of Aggregate Research. *Government and Opposition*, 52(4), pp. 698-722.

Stout, C.T. and Martin, P.J. (2016) Does descriptive representation lead to social desirability bias? Over-reporting of voting among blacks and Latinos in the United States. *Research and Politics*, 3(2), pp. 1-7.

Toma, R. (2020) Dimensiunea reală a emigrației din România. Ce știm și ce nu despre cât de mare e "diaspora" (The real dimension of emigration from Romania. What we know and what we don't know about how big the "diaspora" is). *Libertatea*, February 16. Available at: www.libertatea.ro/stiri/dimensiunea-emigratiei-din-romania-ce-stim-si-ce-nu-despre-cat-de-mare-e-diaspora-2885018 [Accessed on 04 July 2022].