ISSN: 1852-9003 - eISSN: 2660-700X
DOI: https://doi.org/10.14201/rlop.25374

FORECASTING TWO-HORSE RACES IN NEW DEMOCRACIES: ACCURACY, PRECISION AND ERROR

Pronosticando carreras de dos caballos en nuevas democracias: exactitud, precisión y error

Prevendo corridas de dois cavalos em novas democracias: exatidão, precisão e erro

KENNETH BUNKER kabunker@gmail.com [1]

[1] Universidad Diego Portales

Submission: 2020-12-28
Accepted: 2021-10-06
First View: 2022-05-31
Publication: 2022-05-31

Keywords:
Bayesian inference; Election campaigns; New democracies; Public opinion; Plebiscites

Abstract
The purpose of this article is to explore electoral forecasting in two-horse races in new democracies. Specifically, it applies a Bayesian dynamic linear model (coined the Two-Stage Model, TSM) to look at the 2020 Chilean two-question national plebiscite. The ultimate objective is to test the TSM in terms of accuracy (how close the forecast is to the election results), precision (how close the forecast is to other methods of prediction) and error (how the forecast deviates from perfect accuracy/precision). The article finds that while the TSM does appear to be a stable estimator, its accuracy and precision is affected under certain conditions. Using the difference in the results for each of the two questions, the article discusses how sharp and unexpected shifts in electoral preferences can affect forecasts.

Palabras clave: *Inferencia bayesiana; Campañas electorales; Nuevas democracias; Opinión pública; Plebiscitos*

Resumen
El propósito de este artículo es explorar la predicción electoral en carreras de dos caballos en nuevas democracias. Específicamente, aplica un modelo lineal dinámico bayesiano (acuñado el modelo de dos etapas, TSM) para observar el plebiscito nacional de dos preguntas de Chile el 2020. El objetivo final es probar el TSM en términos de exactitud (qué tan cerca está de los resultados de las elecciones), precisión (qué tan cerca está de otros métodos de predicción) y error (qué tanto se desvía de exactitud/precisión perfecta). El artículo encuentra que, si bien el TSM es un estimador estable, su exactitud y precisión se ven afectadas bajo ciertas condiciones. Usando la diferencia en los resultados de las dos preguntas del plebiscito, el artículo discute cómo cambios bruscos e inesperados en las preferencias electorales pueden incidir en los pronósticos.

Palavras-chave: *Inferência Bayesiana; Campanhas eleitorais; Novas democracias; Opinião pública; Plebiscitos*

Resumo
O objetivo deste artigo é explorar a previsão eleitoral em corridas de dois cavalos em novas democracias. Especificamente, ele aplica um modelo linear dinâmico Bayesiano (nomeado de modelo de dois estágios, TSM) para observar o plebiscito nacional de duas perguntas do Chile em 2020. O objetivo final é testar o TSM em termos de precisão (quão próximo está dos resultados?), exatidão (quão próximo está de outros métodos de previsão?) e erro (quanto se desvia da precisão / exatidão perfeita?). O artigo conclui que, embora o TSM seja um estimador estável, sua exatidão e precisão são afetadas sob certas condições. Usando a diferença nos resultados das duas questões do plebiscito, o artigo discute como mudanças repentinas e inesperadas nas preferências eleitorais podem influenciar as previsões.

INTRODUCTION

When it comes to electoral forecasting there is a remarkable lack of research stemming from new democracies in general (Lewis-Beck & Bélanger, 2012; Lewis-Beck & Stegmaier, 2008) and from Latin America in particular (Bunker, 2021; Cantú et al., 2016; Turgeon & Rennó, 2012). While some research has been conducted in the region, it has all focused on presidential elections, leaving both more frequent elections (such as those held at the legislative and municipal levels) and less frequent ones (such as regional referendums and national plebiscites) significantly understudied. However, new advances in forecasting methods (mainly associated to statistical techniques), as well as recent events in countries across the region (more direct, more diverse and more democratic elections), provide the perfect opportunity to advance the understanding of electoral forecasting in new democracies at greater levels of depth.

Gaining greater insight into forecasting is relevant considering the sharp rise of fake news and post-truths surrounding electoral processes (see Allcott & Gentzkow, 2017) in election campaigns across the world (Cassino, 2016). Because new democracies have less safeguards than established ones, and as such are at higher risks of disinformation related vulnerabilities (McKay & Tenove, 2020), it is particularly important to study public opinion trends in their electoral cycles. And because of the rise of direct democracy mechanisms (Altman, 2018), and their implications for governance, it is especially important to study the matter at deeper tiers of citizen electoral engagement beyond its representative scope. This article particularly proposes to look at electoral forecasting in new democracies at the level of national plebiscites—which particularly falls within the two-horse race category (in contrast to multi-candidate or multi-party elections).

In addition to the theoretical warrant, new methodological and computational developments offer a perfect opportunity to apply large-N methods to case studies more efficiently than previously possible. Thus, instead of adopting a traditional custom-fit method purposely tailored to study country-specific dynamics, this article instead applies a previously developed method tested, and to a relevant degree proven to produce accurate and precise forecasts, to a very particular electoral scenario. Specifically, it uses a Bayesian Dynamic Linear Model (DLM) coined the Two-Stage Model (TSM), and applies it to the 2020 Chilean national plebiscite. In this way, this article seeks to contribute not only to the electoral forecasting literature in general, but also the burgeoning body of Chilean electoral studies and public opinion research.

Chile is a particularly suitable case to study electoral forecasting at a more granular level for two major reasons. First, because it is one of the few countries in Latin America that has already accumulated some research on electoral forecasting. As such, this study cannot only contribute to develop a more robust understanding of both election dynamics and public opinion trends in the country, but can also use previous evidence as a point of comparison. The second reason is because the 2020 Chilean national plebiscite was not only a rare event in the institutional history of the country, but was also an election with great political significance (since major constitutional overhaul was on the ballot). In this way, gaining a deeper insight into public opinion trends during a particularly rare and relevant electoral cycle can further contribute to identify the boundaries of accurate and precise electoral forecasting.

The remainder of this article is structured as follows. The following section briefly summarizes some of the main problems related to modern democratic processes and describes how electoral forecasting can contribute to solve some of them. It particularly proposes DLMs in general and the TSM in particular as resolution mechanisms, and pushes the case for the need to advance lines of electoral forecasting research that look at elections other than presidential ones. The third section justifies the case selection (Chile), and describes the electoral process surrounding the 2020 national plebiscite. The fourth section presents the specific research questions and the data, the fifth section shows the main findings, and the final section puts forward a discussion on the greater implications of the results.

FORECASTING ELECTIONS[1]

There is a growing trend of citizens receiving inaccurate information during electoral cycles (Cassino, 2016). This is a problem because voters use their knowledge to inform their decisions (Markus & Converse, 1979). Those with more information are not only more likely to vote (Bartels, 1996; Feddersen & Pesendorfer, 1996; Lassen, 2005; Palfrey & Poole, 1987), but are also more likely to vote for the candidate that yields them with the highest total utility (Ghirardato & Katz, 2006; Matsusaka, 1995). Thus, voters with little or inaccurate information do not only vote less but are also less likely to report having voted for the "right candidate" (Matsusaka, 1995). This can ultimately contribute to the production of artificial, and potentially harmful outcomes for democracy (Fowler & Margolis, 2014; Winters & Weitz-Shapiro, 2013). Because uniformed voters make inefficient assumptions on the distribution of preferences, including their own, they echo preexisting information biases (Nadeau et al., 1993), and ultimately misinform the electoral process (Blais et al., 2009; Larcinese, 2007).

In contrast, democracies that institutionally account for misinformation, and aim at curbing asymmetries, do not only tend to produce elections with higher rates of citizen participation but also tend to produce higher levels of post-election satisfaction (Carpini & Keeter, 1997; Milner, 2002). Thus, curbing information asymmetries is important not only to gap the space between the electoral process and the voter but for the process of democracy itself. Graefe et al. (2014) describe how methods to estimate and relay the "true state" of electoral races date back to at least the early twentieth century. They show how methods have evolved from experts, to polls, to quantitative models, to electronic betting markets. But in the light of some recent and surprising electoral results (such as the UK in 2015 and Australia in 2019), research has moved to attempt to further reduce noise and increase signal.

Data aggregation

Recent research stemming from political and computational science suggests that the solution may lie in poll aggregation (Armstrong et al., 2015; Lewis-Beck &

1. The discussion in this section draws heavily from Bunker (2021).

Dassonneville, 2015; Pasek, 2015; Wang, 2015). Technically, aggregation is simply the combination of data stemming from pre-election information and is grounded in likelihood theories and bracketing principles (see Mannes et al., 2014). Take an election in which two polls are fielded with the intention of predicting the vote difference between the top two candidates as an example. It is likely that the average of the two polls will be a better estimator of the result than any poll chosen at random. Studies have shown that, as a general rule, as more data is considered, accuracy levels improve (Jackson, 2018). And, while experiments and research are still burgeoning, they have already shown that accuracy levels can at least matched those of traditional ones (Graefe et al., 2014).

While aggregation models do not go without limitations, they do provide solutions to many of the pitfalls that traditional methods have not been able to yet solve (Graefe et al., 2015). In contrast to polls and betting markets, they are less vulnerable to late swings and outliers, as they do not generally take potentially biased information at face value. In comparison to quantitative models, they are more versatile, since they can be easily designed to incorporate data from alternative sources. In comparison to experts, they are more likely to tend toward the average preference, because they are naturally more effective in detecting latent trends. And because of their parsimony, they have been on the rise. While their use in media can be traced back to the website FiveThirtyEight, initiatives have since burgeoned (Jackson, 2018). The most basic model is known as the Poll of Polls model.

The Two-Stage Model

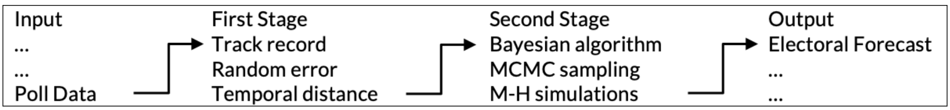
It is in this context that the TSM was developed. The TSM (see Bunker, 2021) essentially uses a poll aggregation method but adds the complexity of space state DLMs (West & Harrison, 1997). Its main objective is to estimate latent trends of support for parties (or candidates) and extrapolate them into the future (Bodell, 2016). Its contribution is its effectivity to combine data over time within the restrictions of probability theory, and to perform real-time tracking of electoral support with minimal and continuous information (Walther, 2015). It is built in the tradition of Jackman, who was among the first to use Bayesian methods to simulate the "true" state of an election using just polls (Jackman, 2005). But it also takes into account the body of literature that has since been developed in North America (Linzer, 2013; Lock & Gelman, 2010; Pickup & Johnston, 2007; Rigdon et al., 2009), the United Kingdom (Fisher & Lewis-Beck, 2015; Hanretty et al., 2016; Whiteley et al., 2016), and continental Europe (Bodell, 2016; Montalvo et al., 2019; Stoetzer et al., 2019; Stoltenberg, 2013; Walther, 2015).

As an extension of these models, the TSM does not intend to be a replacement, yet to propose a simpler set-up, that can be employed in more irregular settings with less specific regulations. For example, in comparison to the Jackman (2005) model, it puts less emphasis on house effects, considering that in developing democracies pollsters tend to be more irregular over time. In comparison to the Linzer (2013) model, the TSM can be easily adapted to any country in which votes are tallied at the national level. All in all, the intention of the TSM is to provide a method of forecasting that bypasses irregularities in the polling industry and electoral system restrictions. It can be applied across a wider number of democracies.

Figure 1 shows a graphical summary of how the TSM works. In the first stage polls are weighed according to three criteria: their accuracy track record, their estimated random error, and their distance from the election. The logic is that polls that are relatively more accurate in one election will be relatively more accurate in the next, polls that structurally anticipate lower levels of random error will be more accurate in comparison to polls that anticipate higher levels of random error, and polls that are fielded closer to the election will be more accurate in comparison to those that are fielded further away from the election. Once these quantities are individually computed, data is normalized to account for different measurements, and each is assigned a specific weight related to their overall corrected expected average error.

In the second stage the weighted polls are used to produce the electoral forecast. Essentially, a Bayesian approach is adopted, in which the parameters are treated as random, but are described by probability distributions. The process begins with the specification of a posterior model, conditional on observed data and prior knowledge (Berger et al., 1988; Bernardo & Smith, 2009). First, it combines the likelihood and prior using the Bayes algorithm (Posterior ∝ Likelihood × Prior) to generate an estimate. Then, it uses a Markov chain Monte Carlo (MCMC) to simulate the election thousands of times. Finally, it simulates the probability of that estimate by means of a Metropolis-Hastings (MH) iterative process.

Figure 1. Summary of the TSM



Source: Author

Consider a vector of polling data *y*, which is assumed to be a sample from a probability model with an unknown parameter vector θ. The objective is to infer its properties. Thus, the model is first represented by a likelihood function where $f(y_i|\theta)$ represents the probability density function. It is then represented by a prior distribution in which θ has a probability distribution p(θ). And because both *y* and θ are random, Bayes theorem can be applied to derive the posterior distribution of θ given data *y*. However, because posteriors often involve multidimensional integrals, they have to be estimated via simulation; normally through MCMC sampling (Gamerman & Lopes, 2006; Tanner & Wong, 1987). Thus, θ at time *t* (with a normal distribution of mean μ and variance σ) is updated until it converges in a posterior (Petris et al., 2009).

Figure 2. Bayesian and MCMC set-up

Bayesian algorithm		MCMC Sampling
$p(\theta\|y) = \dfrac{p(y\|\theta)\,p(\theta)}{p(y)} \rightarrow \theta_t = N(\theta_{t+1}, \sigma)$	$\longrightarrow$	$\theta_t = N(\mu, \sigma)$

Source: Author

To update θ, and move through the chain, this study follows a four-step adaptive Metropolis-Hastings (MH) algorithm (see Metropolis & Ulam, 1949). Ultimately, the objective is to decide if the production of new values of θ are accepted or rejected (Haario et al., 2001; Roberts & Rosenthal, 2009). To explain how this is done, let q(x) be a probability distribution and θ_z the starting state. Then, at each step *t* a proposal state θ_z is generated conditional on the current state. After drawing uniform random numbers, θ_z is accepted or rejected, and updated, according to the previously defined acceptance probability. Figure 3 shows this reiterative process, for *t* = 1,…,T-1.

Figure 3. Metropolis-Hastings set-up

	Step	**Specification**
1.	A proposal state is defined	$\theta_z \sim q(\cdot \mid \theta_{t-1})$
2.	An acceptance probability is set	$\alpha(\theta_z, \theta_{t-1}) = \min[r(\theta_z \mid \theta_{t-1}), 1]$ where $(\theta_z\|\theta_{t-1}) = \frac{p(\theta_z \mid y)\, q(\theta_{t-1} \mid \theta_z)}{p(\theta_{t-1} \mid y)\, q(\theta_z \mid \theta_{t-1})}$
3.	A random number is generated	$\mu \sim \text{Uniform}(0, 1)$
4.	An update is conducted	$\theta_t = \theta_z$ if: $\mu < \alpha(\theta_z \mid \theta_{t-1})$, otherwise: $\theta_t = \theta_{t-1}$

Source: Author

The TSM was tested using data from eleven democracies of the Americas (Bunker, 2021). In that study, the objective in that study was to assess its accuracy in an institutionally unstable setting with relatively low quality of data. The results were remarkable. The TSM produced a more accurate election forecast (in comparison to polls) for 100 percent of the elections (26), and 95 percent of the candidates (126), in the sample. Now, like most models of its nature, the TSM (in its current form) has been only applied at the cross-national level and uniquely for presidential elections. Thus, a contribution to the literature would be to apply the TSM in a different setting. But, not only in a case study where the TSM can be properly tested and compared to itself, but also one in which the context can help provide information on how the preset weights of sample size, method, and distance from the election (described above) can be recalibrated to increase the predictive power of the model. Indeed, it is not only relevant to understand if the TSM is accurate, but also how the TSM can become even more accurate. The next section describes why the 2020 Chilean national plebiscite provides the perfect institutional framework and political context to do just this.

CASE STUDY: CHILE

The selection of Chile as the case study makes sense for a lot of reasons. First, because the TSM was first developed there. Indeed, its origin can be traced back to a study that looked at public opinion trends during the 2013 Chilean presidential election (see Bunker & Bauchowitz, 2016), in which the algorithm worked remarkably well, producing a significantly accurate and precise forecast for all nine candidates that competed in that year's first round of voting. But the selection also makes a lot of sense because the TSM has more recently been tested in three additional presidential elections (2005, 2009, 2017), providing further evidence for a baseline. Together, these studies can help understand if the results of the application of the TSM to the 2020 national plebiscite are in-line with the historical trend or are instead outliers.

In sum, with evidence from Chile, this article proposes to look at a new type of election, which has hitherto been neglected by the literature. It proposes to look at a national plebiscite, that has not only been generally understudied as a generic type of election, but that is also specifically rare and relevant event in the historical context of the Chilean political timeline. Ultimately, this study seeks to not only provide further insight into electoral forecasting at the level of national plebiscites, and two-horse races in general (e.g., run-off elections), using evidence from Chile, but also into electoral forecasting in the context of major elections which are naturally uncertain. This study also seeks to provide a deeper understanding of the particular election at hand, the 2020 Chilean national plebiscite, insofar as it can help identify key moments that took place during the one-year electoral cycle.

The 2020 Chilean national plebiscite

The origin of the 2020 Chilean national plebiscite can be traced back to at least the 2019 social and political crisis, when masses unexpectedly took to the streets to protest against a hike in metropolitan public train (Metro) fares (Sehnbruch & Donoso, 2020). Backed into a corner, the government called for a surprise snap referendum, which at the moment seemed like the only possible solution to diffuse a situation that had suddenly turned critically violent. The objective of the plebiscite was for Chileans to answer two questions: (Q1) "Do you want a New Constitution?", and (Q2) "What type of body should draft the new Constitution?". Each question had two possible answers, or options.[2] While the former simply proposed "Approve" and "Reject" as options, the latter offered "Fully Elected Constitutional Convention" and "Half Elected Constitutional Convention" as options.

The plebiscite was originally planned to take place on the 25 of April of 2020 but was ultimately postponed six months (due to Coronavirus related concerns) and held on the 25 of October of 2020. It was the first major national plebiscite to take place since the 1988 and 1989 referendums, which together marked the transition to democracy in 1990 after nearly seventeen years of dictatorship. Like its predecessors, the 2020 plebiscite was expected to have a long-lasting effect on the party landscape if the results were as bi-modally distributed as those of the 1988 referendum, which asked Chileans if they would like to prolong the dictatorship of Augusto Pinochet or would instead like to transition to democracy (it resulted in 55 percent in favor of the latter). Indeed, if the distribution would have been the case once again in 2020, roughly splitting the country in two, it would have been interpretable as a forecast for a strongly divided country, much like the one in 1990-2020.

The results, however, showed a substantially different picture. Chileans strongly supported change, with 78 percent in favor of drawing a new Constitution and 80 percent in favor of a Fully Elected Constitutional Convention. In other words, in comparison to the 1988 referendum, the 2020 showed a largely unified electorate. But how stable were preferences leading up to the historical event? Did voters make up their mind at the last minute, after being influenced by campaigns or did they decide their votes as soon as the government announced the plebiscite in November of 2019? Furthermore, and more directly related to the matter of this study, were public opinion instruments able to anticipate the results of the election? Did pre-election polls correctly predict that roughly 80 percent of Chileans would vote in one direction? Or did they fail to capture voting intentions in an unexpected, and as such unpredictable, election? Furthermore, was there any way of anticipating the results of the election to a certain degree of accuracy and precision?

2. On the ballot (in Spanish): Q1 was "¿Quiere usted una Nueva Constitución?" and Q2 was "¿Qué tipo de órgano debiera redactar la Nueva Constitución?".

RESEARCH QUESTIONS AND DATA

Linked to electoral forecasting research, presented in Section 2, and Chilean electoral studies and public opinion research, presented in Section 3, this section puts forward three research questions. The first research question is if the TSM can accurately forecast the 2020 Chilean national plebiscite. The answer to this question is relevant because, as mentioned above, it would be some of the first evidence related to forecasting two-horse referendums in new democracies. Also, it would be a test of the stability of the TSM, which has been proved to produce accurate results in multicandidate first round presidential elections but has hitherto been tested in different contexts. To answer this question, the TSM will be applied using the formulae described in Section 2. In terms of data, following the same criteria as in previous studies, all public opinion polls fielded in the election cycle will be included in the dataset.

The second question is if the TSM can produce a more precise forecast than its alternatives. The answer to this question is relevant because within the electoral forecasting body of research there are competing methods, including everything from expert opinions to betting markets to econometric models. And perhaps more importantly, different methods within polls-only models. Hence, to answer this question, the TSM point estimate output will be compared to a series of other methods that can be derived from public opinion polls to produce an election prediction to find out if the TSM output could have provided voters with more precise information than they would have obtained otherwise. Using the same dataset of polls, the TSM point estimate will be compared to each individual pollster's last poll, the average of polls during the campaign, and other methods, such as Lowess and Polynomial specifications.

The third and final question is related to the source of the errors produced by the TSM model. The answer to this question is functional to the specific characteristics of the model and can indicate its stability. Some of the independent variables that will be placed to understand their effect on accuracy will be the time each poll was conducted, the number of people interviewed by each poll, and if the poll was fielded online, face-to-face, via telephone or a mix of any of these methods. To answer this question, this study will simply look at the predictive capacity of each poll, as part of the full set of polls, and in relation to the final result of each of the two questions of the plebiscite. Naturally, the expectation is that polls fielded closer to the plebiscite will produce less error (will be more accurate) and polls that interview more people will produce more error (will be less accurate). Naturally, it will also look at the effect of the campaign and if its interaction with other variables to understand if they had any additional impact on their error.

Table 1 contains a summary of the polls that were used as input for the TSM in order to forecast the 2020 Chilean national plebiscite. All of the polls were published in national mainstream media and collected by the author of this article at the time of their publication. All of the polls met the minimum standards in order to be included into the full sample. This included information regarding fieldwork dates, the number of individuals interviewed, and if they had polled in Chile at any previous point in time. In terms of the latter, and in accordance with the methods of the TSM, all of the polls were assigned a rating bounded between one and zero based on their accuracy in previous elections. In turn, companies that fielded polls for the first time (and as such were unpredictable in terms of their previous record), were assigned a rating equivalent to that of the worst ranked pollster.

Table 1. Summary of the input data

	Full Sample (One year)		Subsample (Three months)	
Pollster	Q1	Q2	Q1	Q2
Activa	22	22	6	5
Cadem	14	14	0	0
CEP	1	1	0	0
CIIR	1	1	1	1
Coes	1	1	0	0
Criteria	8	5	3	2
Data Influye	5	5	2	2
Mori	2	2	1	1
Numen	2	1	2	1
StatKnows	2	2	1	1
TOTAL	58	54	16	13

Source: Author with data from each individual pollster.

In summary, Table 1 distinguishes between a full sample (polls fielded between the 16 of November of 2019 and election night) and a subsample (polls fielded between the 25 of July of 2020 and election night). It is important to note that public opinion polls that contain information on voter preferences (that can be interpreted as voting intentions) can legally only be published until fifteen days before the election. As such, the latest poll registered in the full sample was fielded before the 11 of October of 2020 (to be precise on the 9 of October of 2020). The data presented in the Table suggests at least three things. First, that there were more polls fielded for the first of the two questions. Second, that one company (Activa) polled significantly more than all other companies. And third, that only around one fourth of all polls were fielded during the three-month election campaign cycle.

Table 2 shows a summary of poll predictions. Here it is important to note that the percentages reflect the average prediction made by each pollster. If a pollster only fielded one poll, the percentage is equal to that poll's prediction. It is also important to note that within each poll, the estimations do not necessarily add up to one hundred, since the sample does not consider likely voter models, and most polls also include estimates relative to interviewees that answered "don't know" or had "no opinion" to the questions.[3] Finally, it is important to note that not all posters fielded polls during the campaign cycle, as visible in Table 1, but among those who did, the trend does not show any major deviations. In other words, and considering that most polls were fielded in the nine months previous to the

Table 2. Summary of public opinion results

	Full Sample (One year)		Subsample (Three months)	
	Q1	**Q2**	**Q1**	**Q2**
Activa	71.8	12.5	70.6	12.4
Cadem	70.0	22.4	--	--
CEP	77.0	13.0	--	--
CIIR	75.0	12.0	75.0	12.0
Coes	85.5	8.1	--	--
Criteria	72.4	18.1	73.0	18.0
Data Influye	72.8	17.2	71.0	15.5
Mori	67.0	16.0	66.0	15.0
Numen	40.2	34.1	40.2	34.1
StatKnows	55.5	38.9	55.4	43.1

Source: Author with data from each individual pollster.

3. The method does not consider likely voter polls too produce a wider range of results and, as such, increase the emphasis of the "let the data speak for itself" approach. Also, including likely voter models which are essentially different across pollsters, could risk introducing unexpected bias. At any rate, very few polls actually conduct likely voter polls, as Table 3 shows.

election, the Table suggests that electoral preferences generally tended to be stable during the year leading up to the plebiscite.

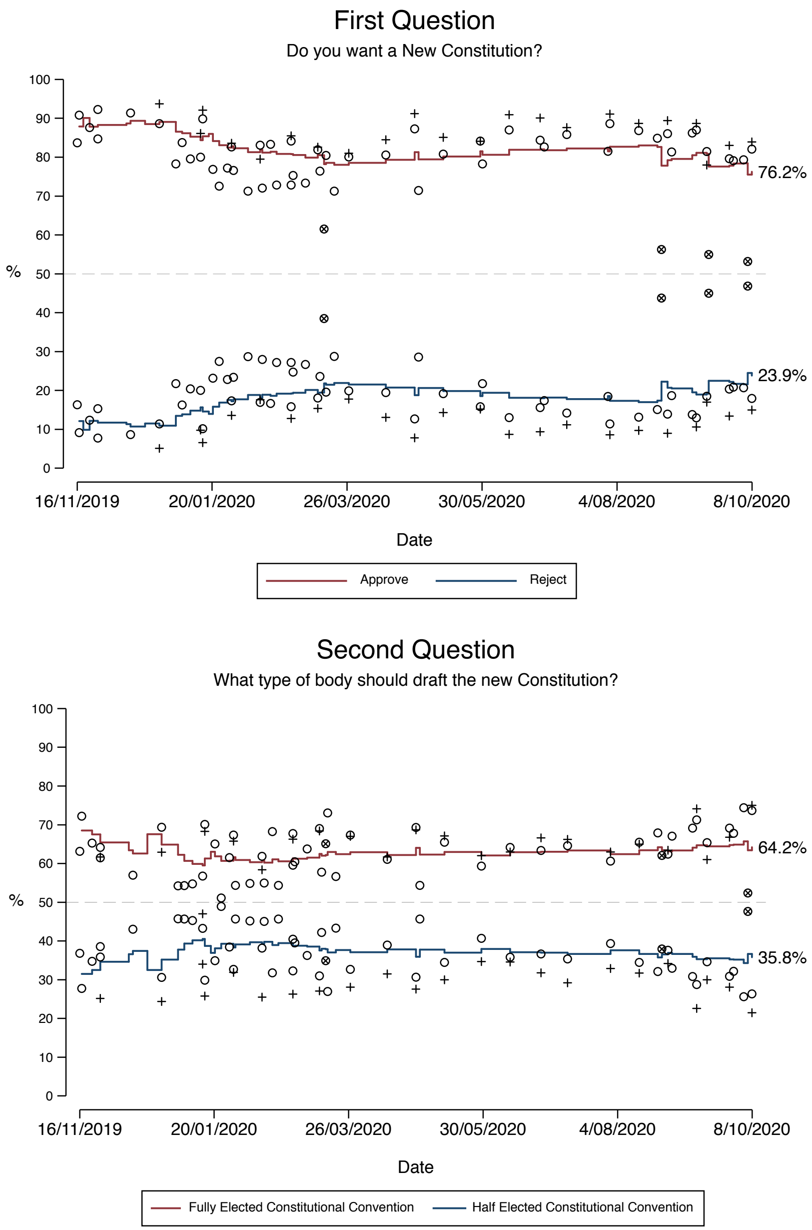
Now, before moving on to the findings, it is relevant to double-check the overall distribution of the polls to avoid including biased data into the sample. A simple inspection immediately suggests that two of the pollsters in the full sample stand out because of their odd deviations: Numen and StatKnows. Both of these companies, which incidentally are the only two non-Chilean firms, produced significantly lower differentials than all other pollsters; they were strongly biased in favor of the "Reject" and "Half Elected Constitutional Convention" options. For example, in the first question, as most polls showed an average advantage of 60 percent in favor of the first option, both Numen and StatKnows showed an average advantage of less than 15 percent. Because both of these companies were new to the Chilean polling industry, they were structurally assigned the lowest rating in the sample yet included all the same. However, because of their remarkably odd deviations, and strong outlier behavior, their influence on the final forecast will be studied with greater detail in the following section.

FINDINGS

The first research question is if the TSM can produce equally accurate results for the 2020 national plebiscite as it has for first round multiparty presidential elections. As the question suggests, there are two parts to it. First, if the TSM can produce an accurate result for the plebiscite, and second, if those results are more accurate than what the TSM has previously produced. Beginning with the former, Figure 4 shows the TSM forecast (line) superimposed over the polls (circles). Note that the polls have been re-scaled to sum 100 percent, as this is part of the essential transformations proposed by the TSM in the first stage of computations. Likely voter polls (crosses) are shown but not used to produce the forecast. The plot on the left shows the trend for the first question and the plot on the right shows the trend for the second question.

In terms of the first of the two questions, the TSM was remarkably accurate. While the actual result for the "Approve" option was 78.3 percent, the TSM forecast was 76.2 percent. In other words, the TSM forecast produced an absolute error of 2.1 percent. In terms of the second of the two questions, the TSM was significantly off its mark. In this case, while the actual result for the "Fully Elected Constitutional Convention" option was 79 percent, the TSM forecast was 64.2 percent. Ergo, the TSM forecast produced an absolute error of 14.8 percent. These mixed results are indeed alarming and worth inspecting at greater depth.

Figure 4. TSM Forecast and Rescaled Polls



Source: Author. *Note*: Circles fit with crosses are Numen and StatKnows polls.

Is it that the TSM is structurally accurate (as evidenced by the first question) and the error of the second question can be explained contextually, or is it that the TSM is imprecise (as evidenced by the second question) and the accuracy of the first question is an exception to the norm? Naturally, the former would be in line with the expectations. But to prove that this is the case there should be clear evidence that the errors are neither election-specific (i.e., the TSM cannot produce an accurate forecast for plebiscites), nor country-specific (i.e., the TSM cannot produce an accurate forecast for Chile). The question, then, is a matter of consistency. Which of the forecasts for the plebiscite is consistent with the baseline? If the TSM forecast for the first question is in line with the TSM forecast for other types of elections, and for previous elections in Chile, then the outlier is clearly the forecast for the second question.

So, let us begin with the former: is the forecast of the TSM for the first question of the plebiscite in line with the TSM forecast for other types of elections? In its application to twenty six first round presidential elections in eleven countries, Bunker (2021) shows that the mean absolute error of the TSM is significantly lower than the mean absolute error of the polls. More specifically, it shows that while the model erred by an average of 4.1 percent, the polls erred by an average of 5.2 percent. Which of the observations in this study are in line with that pattern? The answer is that the error associated to the first question is more consistent with the average error produced by the TSM elsewhere (in other types of elections) than the error associated to the second question. While the absolute mean difference between the TSM error elsewhere and the TSM error for the first question is 2 percent, the absolute difference between the TSM error elsewhere and the TSM error for the second question is 10.7 percent. In this way, there is no evidence that the second question fits the pattern. Instead, there seems to be evidence that the error is instead correlated to the specific question.

Now, we turn to the latter: is the forecast of the TSM for the first question of the plebiscite in line with the TSM forecast for previous forecasts of elections in Chile? In its application to four presidential elections in Chile (2005, 2009, 2013 and 2017), the mean absolute error of the TSM is significantly lower than three percent (see Bunker & Bauchowitz, 2016). Which of the observations in this study are in line with that pattern? The answer, again, is that the average error associated to the first question is more consistent with the average error produced by the TSM in Chile (in previous elections) than the error associated to the second question. While the absolute difference between the TSM error in previous elections and the TSM error for the first question is 1.6 percent, the absolute difference between the TSM error in previous elections and the TSM error for the second question is 11.8 percent. Once again, there is no evidence that the forecast for the second question fits the pattern; it seems that the error is instead correlated to the specific question.

In sum, there is no evidence that the error is related to the model (TSM), the type of election (plebiscite) or the country (Chile). Instead, it seems to be associated to the second question itself, which seems to have been particular in more than one way. In retrospect, there are several observations that buttress that idea. Indeed, few election observers would dispute the fact that the first question ("Do you want a New Constitution?") was significantly more popular than the latter ("What type of body should draft the new Constitution?"). There is some evidence that supports this claim. For example, the proportion of people who did not answer voting intention questions was significantly higher in the latter question. In the first question, the average "does not know/no opinion" response was 8.1 percent for the full sample and 10.1 percent for the subsample. In the second question, the average "does not know/no opinion" response was 11.3 percent for the full sample and 12.9 percent for the subsample. This suggests that voters were more decided and likely informed for the first question than for the second. Also consistent with the claim that the attention surrounding the first of the two questions was more prevalent, is the fact that campaign contributions were significantly lopsided in its favor. Indeed, of the total 520 million Chilean pesos donated to the campaigns, 484 million (93 percent) went to the first question as just thirty six million (7 percent) went to the second question (Servel, 2020).

All in all, it seems that the second question was very particular in its nature. First, because it was the less defining question of the two questions asked. Indeed, if people would have rejected the first question, then the second question would not have mattered. Second, as polls show, less people were actually aware of what the second question was before the election. This was confirmed by election voting patterns: while the total turnout was equal for both questions (around 7.5 million votes), Q1 totaled less than 40 thousand invalid votes (blank and null votes), Q2 totaled over 400 thousand of the same. Of course, the lopsided campaign financing in favor of the first question did not help the second question. Indeed, anecdotal evidence suggests that TV campaign spots (broadcasted twice a day for the final 30 days of the campaign) significantly favored the first question over the second one.

The second research question is if the TSM adds any value to the information we could have obtained otherwise to anticipate the results of the election. The traditional manner to answer this question is to compare the results with its alternatives. In the case of the TSM, or DLMs that use poll-only data in general, the standard comparison is to both other methods of the like as well as to the polls themselves. While the comparison of poll-only DLMs to the same polls they use as input is not necessarily a fair comparison, since models are structurally built to bracket results and produce averages, it is still a relevant comparison if the objective is to know if the voters could have had access to better information during electoral cycles. In this way, it is important to show the comparison of the TSM to other methods, as well as the polls, in different combinations and configurations.

Table 3 shows the final poll of each pollster that fielded a poll during the three-month campaign cycle and its error in comparison to the result of the election. The parameter of interest is the result for the "Approve" and the "Fully Elected Constitutional Convention" options. The data suggests that the last polls fielded in the cycle overestimated the outcome of the first question and underestimated the outcome of the second question. Two observations are worth noting. First, that the two polls mentioned above as possible outliers (Numen and StatKnows) were indeed off by more than twenty percent on average, preliminarily suggesting that the model does better without them. The other observation is on the uncertainty surrounding the second question. Indeed, as two polls fell within the traditional three percent margin of error for the first question, just one did the same for the second question. Errors for the second question were also consistently and significantly high.

Table 3. Poll Predictions and TSM Forecast

	First Question, Winning Option (78.3%)		Second Question, Winning Option (79%)	
Polls	**Prediction**	**Error**	**Prediction**	**Error**
Activa*	83.9	5.6	75.0	4.0
CIIR	75.0	-3.3	65.0	14.0
Criteria*	72.0	-6.3	59.0	20.0
Data Influye	69.0	-9.3	61.0	18.0
Mori	78.0	-0.3	61.0	18.0
Numen	38.5	-39.8	38.5	40.5
StatKnows	55.4	-22.9	54.0	25.0
Methods	**Forecast**	**Error**	**Forecast**	**Error**
TSM	76.2	2.1	64.2	14.8
TSM 2.0.	81.6	3.3	66.4	12.6
30-day average	69.4	8.9	59.3	19.7
Lowess	73.7	4.6	67.9	12.1
Lpoly	72.1	6.2	67.0	13.0

Source: Author. *Likely voter models

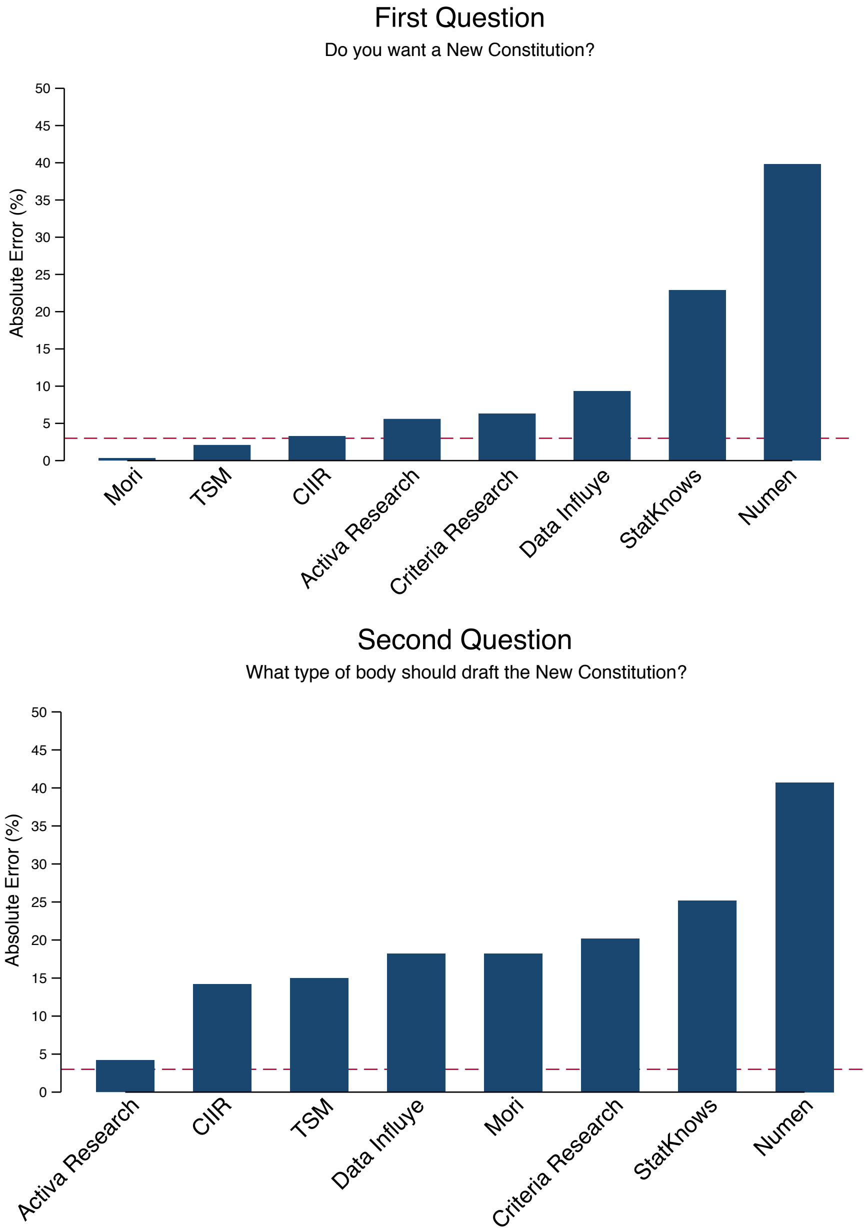
Table 3 also shows the results of the TSM and some other methods used to forecast election results from polls. It considers the full sample of polls. As anticipated above, it shows that the TSM forecast was 76.2 percent, and its associated error was 2.1 percent. This is simply the point estimate for the trend shown in Figure 4. It shows that in comparison to polls, the TSM was more precise. In both questions, the TSM forecast would have given more information than any poll chosen at random—notwithstanding its significant error for the second question. TSM 2.0. shows the same process, but excludes Numen and StatKnows from the sample, because of their outlier behavior. It suggests that while the forecast would have been worst in case of the first question, it would have been better in case of the second. At any rate, it does not generate major differences in the overall order of precision considering both the TSM and the polls.

In addition to the TSM, the Table 3 shows that other common methods used to aggregate polls, such as Lowess and Polynomial specifications, would not have been more precise than the TSM in the first question, yet would have been marginally more accurate for the second question. Because both methods are more sensitive to outliers, placing less weight on historical trends, they are structurally able to shift their forecast at the last minute. In a way, this can be interpreted as evidence that a shift in preferences took place at the end of the campaign for the latter of the two questions.[4] This idea will be further explored in the following subsection, as it may be able to explain the error. At any rate, it is worth noting that a simple average of the last 30 days (a common benchmark) would have been significantly off the mark.

Figure 5 shows the errors of all pollsters for the first and second questions. The superimposed red line shows the three percent reference marker, which is what polls normally use as their standard margins of error. The plot on the right shows that the TSM placed second, only after Mori, with an error of 2.1 percent. The plot on the left shows that the TSM placed third, only after Activa and CIIR, with an error of 14.8 percent. As can be inferred from the data above, the error related to the second question was significantly higher, and likely related to the specific context of that particular question (and not to the model, the election, or the country). Indeed, with the exception of Activa (which fielded more polls than any other pollster, and also fielded the last poll in the cycle) that produced an error of only

4. The Lowess specification was run considering Cleveland's tricube weighting function with a bandwidth of 0.8. The Lpoly kernel was run considering the default Kernel Epanechnikov using a rule-of-thumb estimator. Both are based on the specified literature and are the default and standard measures for the statistical package Stata v. 15. These two measures are thus able to capture more natural variations in the data than the TSM, which is set to look for more structural variations. Methodologically, the Lowess and Lpoly are found to be useful estimators, but if used with default settings, will produce inconsistent results.

Figure 5. TSM Forecast and Poll Prediction Errors



Source: Author.

one percent, all pollsters were significantly off the mark, most notably Numen and Statknows. Even if these polls were rescaled to 100 percent, to represent the proportions of the valid options, they still would have been off by around 20 percent. In sum, the TSM produced the most accurate average forecast for the election.

The third and final research question is related to the errors produced by the polls fielded to predict the results of the 2020 Chilean national plebiscite. Here the objective is to understand which variables were more relevant in determining accuracy levels. Previous studies that have looked at large-N cross-national data have found the number of pollsters included in the dataset have the strongest effect on error; they have also found that elections that are contested under more restrictive rules (when the president is elected by simple majority) produce lower errors than elections contested under more permissive ones (when the president is elected by absolute majority) (see Bunker, 2021). However, because of the nature of this particular article (a case study), it is difficult to understand if those variables hold significant beyond what could be expected theoretically for an election with few pollsters and only two answers to each question. However, they do serve as theoretical vectors, and open up avenues to further explore sources of accuracy within single system elections.

Thus, to explore the nature of the difference between predictions and results, the remainder of this section looks at the specific characteristics of the data. For this, the outcome variable is the absolute difference between the prediction for the winning option ("Approve" in the case of the first question, and "Fully Elected Constitutional Convention" in the case of the second question) and their respective results (78.3 percent for the former, and 79 percent for the latter). The independent variables are the number of interviewees in each poll (N), a dummy indicating if the polls prediction was the result of a likely voter model (LVM), the percentage of interviewees that answered "don't know" or had "no opinion" (DK/NO), a dummy indicating if the poll was conducted online or via any other method (Online), and a log transformation of Delta, to capture the effect of the interaction between the three month campaign and the number of days between the poll and election night.

The following table shows several models referring to the sources of error in the first and second questions of the 2020 Chilean national plebiscite. The first and third models are simply baselines, containing a linear regression with robust standard errors. The second and fourth models are the same as the baselines, but cluster the data by pollster to control for house effects. While the pair-wise difference in the beta coefficients does not vary, the standard errors do, revealing significance patterns. Several other specifications were explored, such as one that included an interaction term between LVM polls and the campaign period, and one that included an interaction term between Delta and the campaign period. The models below were chosen do to their methodological simplicity and consistency.

In terms of the first question, Table 4 shows that while sample size appears to be significant, the beta coefficient is too low to have any meaningful effect on the results. At any rate, there is evidence to infer that as the proportion of DK/NO respondents increases, the error also increases. Now, what is even more interesting, is that likely voter models were less accurate than regular polls. In other words, LVM polls produced larger errors than non-LVM polls. This is already some evidence of the stability of electoral preferences in the first question. But it is not the only evidence. Indeed, the same can be deduced from the idea that there was no significant difference between online, face-to-face, phone, or mixed methods polls. They were all equally accurate. But most importantly, there is no evidence to suggest that time had a significant impact on accuracy. Because the logged Delta is not significant, there is no indication that time influenced accuracy. These findings are

Table 4. Sources of Error

	First Question				Second Question			
	M1 (Baseline)		M2 (House effect)		M3 (Baseline)		M4 (House effect)	
	beta	rse	beta	rse	beta	rse	beta	rse
Interviews (N)	0.002***	0.00	0.002***	0.00	0.000***	0.00	0.000**	0.00
Likely Voter (LVM)	3.037*	1.54	3.037***	1.17	-5.470***	1.63	-5.470***	1.49
DK/NO	0.203	0.13	0.203***	0.08	0.658***	0.16	0.658**	0.22
Online	-0.345	1.19	-0.345	1.34	1.848	1.96	1.848	3.01
Log(Delta)	0.992	0.61	0.992	0.76	2.461***	0.70	2.461***	0.53
constant	-1.516	3.50	-1.516	3.21	5.431	4.54	5.431	3.74
N	80		80		76		76	
Clusters			10				10	
F	47.22		259.57		67.63		490.26	
R-squared	0.636		0.636		0.709		0.709	
Root MSE	4.072		4.072		4.447		4.447	

Dependent variable: Absolute difference between winning option final result and poll prediction. Note: ***: p > 0.1; **: p ≤ 0.05; *: p ≤ 0.1.

Source: Author.

consistent with the literature, particularly the work of Jennings et al. (2020), which shows that polls are generally accurate and informative early in campaign cycles.

In terms of the second question, there is some evidence that echoes the preliminary findings above; for example, that sample size was largely irrelevant to determine accuracy and that polls with higher levels of DK/NO respondents showed higher levels of error. And, as above, the method was irrelevant—all polls were equally inaccurate.

Remarkably, however, there is evidence that in the second question accuracy levels increased over time, even in the context of its higher baseline error (as interpretable by the constant). This is similar to the above, and consistent with the idea that as polls are fielded closer to the election, their accuracy tends to increase (Jennings & Wlezien, 2016); even though the accuracy sweet spot generally tends to fall weeks out from election night (Jennings et al., 2020). At any rate, it is relevant to note that, in contrast to the first question, time did play a role in the accuracy of the second question, suggesting that the political context did have an effect that was invisible, or absent, surrounding the first question. Yet, that is not all. There is also other evidence that seems to indicate that preferences in the second question were more volatile than preferences in the first question, as hypothesized above. For example, while regular polls produced lower errors than LVM polls in the first question, LVM polls produced lower errors than regular polls in the second question. The relationship is inverse, strong, and significant. In essence, the evidence shows that more sophisticated methods were necessary to grasp the state of the race in the second question—preferences were more disperse, and as such less predictable.

DISCUSSION

The purpose of this article has been to test a Bayesian Dynamic Linear Model (DLM) developed to forecast multiparty elections in a new, rare and understudied context. It specifically applied the Two-Stage Model (TSM) to the 2020 Chilean two-question national plebiscite. In doing so it has provided some of the first evidence of electoral forecasting for plebiscites in new democracies. It has also contributed by providing a plausible account of public opinion trends that will be useful for Chilean electoral studies and public opinion research. In sum, this study has shown that the TSM can produce an accurate (an absolute measure) and precise (a relative measure) forecast for two-horse races, and that the sources of error related to the model are correlated to structural features of polls, such as the number of interviews conducted, the proportion of valid responses and the method through which polls are conducted.

Interestingly, as the forecast for one of the two questions was remarkably accurate (Q1), the forecast for the other was significantly off its mark (Q2). In comparison to evidence at the regional- (eleven countries in Latin America) and country- (four presidential elections in Chile) levels, it showed that the outlier was clearly the second question. In other words, that the error was neither model-specific, election-specific, or country-specific, but question-specific. This idea was reinforced after looking at a battery of determinants of error in both questions. Because the evidence pointed to the fact that there were irregular shifts in preferences for the second question, and there is a particular law which does not allow for polls to take place the final fifteen days before election night, the TSM could not capture late shifts. The lesson here is that the weight of polls fielded late in the campaign should be increased when there are irregular patterns in the electoral cycle, as evidence from the application of the Lowess method shows. Recalibration to consider LVM polls at a higher weight in these irregular, unstable scenarios should also be considered.

At the theoretical level this study is a contribution to the literature since it advances the understanding of forecasting two-horse races in Latin America. Because many of the countries in the region use two-round elections to choose their leaders, the methods applied here can be easily fit to forecast presidential runoffs. And because the error related to the second question can be, at least partially explained because of the volatility related to the second question and the ban on polls, the results should be more accurate in more stable elections. At the national level, this article has contributed to Chilean electoral studies and public opinion research by providing evidence of trends during the 2020 Chilean national plebiscite, an election that will go down in the books as a crucial moment in the country's political and constitutional history. At the same time, it has contributed to identify some crucial questions that will also help further explain the critical juncture. Most importantly, why was there a surge in electoral preferences in favor of the winning option of the second question at the last moment of the campaign cycle?

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Altman, D. (2018). Latin America. In M. Qvortrup (Ed.), *Referendums Around the World: With a Foreword by Sir David Butler* (pp. 185–211). Springer International Publishing. https://doi.org/10.1007/978-3-319-57798-2_6

Armstrong, J., Green, K., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, *68*(8), 1717–1731. https://doi.org/10.1016/j.jbusres.2015.03.031

Bartels, L. M. (1996). Uninformed Votes: Information Effects in Presidential Elections. *American Journal of Political Science*, *40*(1), 194–230. https://doi.org/10.2307/2111700

Berger, J. O., Wolpert, R. L., Bayarri, M. J., DeGroot, M. H., Hill, B. M., Lane, D. A., & LeCam, L. (1988). The Likelihood Principle. *Lecture Notes-Monograph Series*, *6*, iii-v+vii-xii+1-199.

Bernardo, J. M., & Smith, A. F. M. (2009). *Bayesian Theory*. John Wiley & Sons.

Blais, A., Gidengil, E., Fournier, P., & Nevitte, N. (2009). Information, visibility and elections: Why electoral outcomes differ when voters are better informed. *European Journal of Political Research*, *48*(2), 256–280. https://doi.org/10.1111/j.1475-6765.2008.00835.x

Bodell, M. H. (2016). *Bayesian poll of polls for multi-party systems*. [Master's thesis, Linköping University]. https://liu.diva-portal.org/smash/get/diva2:945786/FULLTEXT01.pdf

Bunker, K. (2021). A two-stage model to forecast elections in new democracies. *International Journal of Forecasting*, *36*(4), 1407–1419. https://doi.org/10.1016/j.ijforecast.2020.02.004

Bunker, K., & Bauchowitz, S. (2016). Electoral Forecasting and Public Opinion Tracking in Latin America: An Application to Chile. *Política*, *54*(2), 207–233. http://www.redalyc.org/articulo.oa?id=64551061008

Cantú, F., Hoyo, V., & Morales, M. A. (2016). The Utility of Unpacking Survey Bias in Multiparty Elections: Mexican Polling Firms in the 2006 and 2012 Presidential Elections. *International Journal of Public Opinion Research*, *28*(1), 96–116. https://doi.org/10.1093/ijpor/edv004

Carpini, M. X. D., & Keeter, S. (1997). *What Americans Know about Politics and Why It Matters* (1st ed.). Yale University Press.

Cassino, D. (2016). Why Pollsters Were Completely and Utterly Wrong. *Harvard Business Review*, *9*. https://hbr.org/2016/11/why-pollsters-were-completely-and-utterly-wrong

Feddersen, T. J., & Pesendorfer, W. (1996). The Swing Voter's Curse. *The American Economic Review*, *86*(3), 408–424. https://www.kellogg.northwestern.edu/faculty/fedderse/homepage/papers/curse.pdf

Fisher, S. D., & Lewis-Beck, M. S. (2015). Forecasting the 2015 British general election: The 1992 debacle all over again? *Electoral Studies*, *41*. https://doi.org/10.1016/j.electstud.2015.11.016

Fowler, A., & Margolis, M. (2014). The political consequences of uninformed voters. *Electoral Studies: An International Journal*, *34*, 100–110. https://doi.org/10.1016/j.electstud.2013.09.009

Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2nd ed.). CRC Press.

Ghirardato, P., & Katz, J. N. (2006). Indecision Theory: Weight of Evidence and Voting Behavior. *Journal of Public Economic Theory*, *8*(3), 379–399. https://doi.org/10.1111/j.1467-9779.2006.00269.x

Graefe, A., Armstrong, J. S., Jones, R. J., & Cuzán, A. G. (2014). Combining forecasts: An application to elections. *International Journal of Forecasting*, *30*(1), 43–54. https://doi.org/10.1016/j.ijforecast.2013.02.005

Graefe, A., Küchenhoff, H., Stierle, V., & Riedl, B. (2015). Limitations of Ensemble Bayesian Model Averaging for forecasting social science problems. *International Journal of Forecasting*, *31*(3), 943-951. https://www.sciencedirect.com/science/article/pii/S0169207014001769

Haario, H., Saksman, E., & Tamminen, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli*, *7*(2), 223–242. https://doi.org/10.2307/3318737

Hanretty, C., Lauderdale, B., & Vivyan, N. (2016). Combining national and constituency polling for forecasting. *Electoral Studies*, *41*, 239–243. https://doi.org/10.1016/j.electstud.2015.11.019

Jackman, S. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, *40*(4), 499–517. https://doi.org/10.1080/10361140500302472

Jackson, N. (2018). The Rise of Poll Aggregation and Election Forecasting. In L. R. Atkeson & M. Alvarez (Eds.), *The Oxford Handbook of Polling and Survey Methods*. https://doi.org/10.1093/oxfordhb/9780190213299.013.28

Jennings, W., Lewis-Beck, M., & Wlezien, C. (2020). Election forecasting: Too far out? *International Journal of Forecasting*, *36*(3), 949–962. https://doi.org/10.1016/j.ijforecast.2019.12.002

Jennings, W., & Wlezien, C. (2016). The Timeline of Elections: A Comparative Perspective. *American Journal of Political Science*, *60*(1), 219–233. https://www.jstor.org/stable/24583060

Larcinese, V. (2007). Does political knowledge increase turnout? Evidence from the 1997 British general election. *Public Choice*, *131*, 387–411. https://doi.org/10.1007/s11127-006-9122-0

Lassen, D. D. (2005). The Effect of Information on Voter Turnout: Evidence from a Natural Experiment. *American Journal of Political Science*, *49*(1), 103–118. https://doi.org/10.2307/3647716

Lewis-Beck, M. S., & Bélanger, É. (2012). Election forecasting in neglected democracies: An introduction. *International Journal of Forecasting*, *28*(4), 767–768. https://doi.org/10.1016/j.ijforecast.2012.04.006

Lewis-Beck, M. S., & Dassonneville, R. (2015). Forecasting elections in Europe: Synthetic models. *Research & Politics*, *2*(1), 1–11. 2053168014565128. https://doi.org/10.1177/2053168014565128

Lewis-Beck, M. S., & Stegmaier, M. (2008). The Economic Vote in Transitional Democracies. *Journal of Elections, Public Opinion and Parties*, *18*(3), 303–323. https://doi.org/10.1080/17457280802227710

Linzer, D. A. (2013). Dynamic Bayesian Forecasting of Presidential Elections in the States. *Journal of the American Statistical Association*, *108*(501), 124–134. https://doi.org/10.1080/01621459.2012.737735

Lock, K., & Gelman, A. (2010). Bayesian Combination of State Polls and Election Forecasts. *Political Analysis*, *18*(3), 337–348. https://doi.org/10.1093/pan/mpq002

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, *107*(2), 276-299. https://doi.org/10.1037/a0036677

Markus, G. B., & Converse, P. E. (1979). A Dynamic Simultaneous Equation Model of Electoral Choice. *American Political Science Review*, *73*(4), 1055–1070. https://doi.org/10.2307/1953989

Matsusaka, J. (1995). Fiscal Effects of the Voter Initiative: Evidence from the Last 30 Years. *Journal of Political Economy*, *103*(3), 587–623. http://www.jstor.org/stable/2138700

McKay, S., & Tenove, C. (2020). Disinformation as a Threat to Deliberative Democracy. *Political Research Quarterly*, *74*(3), 703-717. https://doi.org/10.1177/1065912920938143

Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, *44*(247), 335–341. https://doi.org/10.1080/01621459.1949.10483310

Milner, H. (2002). *Civic Literacy: How Informed Citizens Make Democracy Work* (1st ed.). Tufts.

Montalvo, J. G., Papaspiliopoulos, O., & Stumpf-Fétizon, T. (2019). Bayesian forecasting of electoral outcomes with new parties' competition. *European Journal of Political Economy*, *59*, 52–70. https://doi.org/10.1016/j.ejpoleco.2019.01.006

Nadeau, R., Cloutier, E., & Guay, J.-H. (1993). New Evidence about the Existence of a Bandwagon Effect in the Opinion Formation Process. *International Political Science Review / Revue Internationale de Science Politique*, *14*(2), 203–213. https://www.jstor.org/stable/1601152

Palfrey, T. R., & Poole, K. T. (1987). The Relationship between Information, Ideology, and Voting Behavior. *American Journal of Political Science*, *31*(3), 511–530. https://doi.org/10.2307/2111281

Pasek, J. (2015). Predicting Elections: Considering Tools to Pool the Polls. *Public Opinion Quarterly*, 79(2), 594-619. https://doi.org/10.1093/POQ/NFU060

Petris, G., Petrone, S., & Campagnoli, P. (2009). Dynamic linear models. In P. Campagnoli, S. Petrone, & G. Petris (Eds.), *Dynamic Linear Models with R* (pp. 31–84). Springer. https://doi.org/10.1007/b135794_2

Pickup, M., & Johnston, R. (2007). Campaign trial heats as electoral information: Evidence from the 2004 and 2006 Canadian federal elections. *Electoral Studies: An International Journal*, *26*(2), 460–476. https://doi.org/10.1016/j.electstud.2007.03.001

Rigdon, S. E., Jacobson, S. H., Cho, W. K. T., Sewell, E. C., & Rigdon, C. J. (2009). A Bayesian prediction model for the U.S. presidential election. *American Politics Research*, *37*(4), 700–724. https://doi.org/10.1177/1532673X08330670

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367. https://doi.org/10.1198/jcgs.2009.06134

Sehnbruch, K., & Donoso, S. (2020). Social Protests in Chile: Inequalities and other Inconvenient Truths about Latin America's Poster Child. *Global Labour Journal*, *11*(1). https://doi.org/10.15173/glj.v11i1.4217

Servel. (2020). *Gasto y Propaganda Electoral Plebiscito Nacional 2020*. Servel. https://www.plebiscitonacional2020.cl/propaganda-electoral-plebiscito-nacional-2020/

Stoetzer, L. F., Neunhoeffer, M., Gschwend, T., Munzert, S., & Sternberg, S. (2019). Forecasting Elections in Multiparty Systems: A Bayesian Approach Combining Polls and Fundamentals. *Political Analysis*, *27*(2), 255–262. https://doi.org/10.1017/pan.2018.49

Stoltenberg, E. A. (2013). *Bayesian Forecasting of Election Results in Multiparty Systems*. [Master's thesis, University of Oslo]. http://urn.nb.no/URN:NBN:no-37414

Tanner, M. A., & Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.2307/2289457

Turgeon, M., & Rennó, L. (2012). Forecasting Brazilian presidential elections: Solving the N problem. *International Journal of Forecasting*, *28*(4), 804–812. https://doi.org/10.1016/j.ijforecast.2012.04.003

Walther, D. (2015). Picking the winner(s): Forecasting elections in multiparty systems. *Electoral Studies*, *40*, 1–13. https://doi.org/10.1016/j.electstud.2015.06.003

Wang, S. S.-H. (2015). Origins of Presidential poll aggregation: A perspective from 2004 to 2012. *International Journal of Forecasting*, *31*(3), 898–909. https://doi.org/10.1016/j.ijforecast.2015.01.003

West, M., & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/b98971

Whiteley, P., Clarke, H. D., Sanders, D., & Stewart, M. C. (2016). Forecasting the 2015 British general election: The Seats-Votes model. *Electoral Studies*, *41*, 269–273. https://doi.org/10.1016/j.electstud.2015.11.015

Winters, M. S., & Weitz-Shapiro, R. (2013). Lacking Information or Condoning Corruption: When Do Voters Support Corrupt Politicians? *Comparative Politics*, *45*(4), 418–436. https://www.jstor.org/stable/43664074