International Journal of Public Opinion Research Vol. 28 No. 1 2016

doi:10.1093/ijpor/edv004 Advance Access publication 3 April 2015

The Utility of Unpacking Survey Bias in Multiparty Elections: Mexican Polling Firms in the 2006 and 2012 Presidential Elections

Francisco Cantú[1], Verónica Hoyo[2] and Marco A. Morales[3]

[1]University of Houston, USA; [2]UC San Diego, La Jolla, CA, USA; [3]Instituto Tecnológico Autónomo de México, San Ángel, México

Abstract

This article proposes a more nuanced method to assess the accuracy of preelection polls in competitive multiparty elections. Relying on data from the 2006 and 2012 presidential campaigns in Mexico, we illustrate some shortcomings of commonly used statistics to assess survey bias when applied to multiparty elections. We propose the use of a Kalman filter-based method that uses all available information *throughout* an electoral campaign to determine the *systematic error* in the estimates produced for each candidate by all polling firms. We show that clearly distinguishing between sampling and systematic biases is a requirement for a robust evaluation of polling firm performance, and that house effects need not be unidirectional within a firm's estimates or across firms.

As recent examples across the region show, the use and the value of preelection polls is still debated across Latin America. In El Salvador, although preelectoral surveys for the 2014 presidential run-off election predicted an overwhelming victory for the candidate of the *Alianza Republicana Nacionalista*, Norman Quijano, the official result gave Quijano a marginal difference of <0.5%.[1] In Costa Rica, the winner of the first round of the 2014 presidential elections was Luis Guillermo Ortíz, a candidate who systematically appeared as third in the polls, well behind the candidates of the

All correspondence concerning this article should be addressed to Francisco Cantú, University of Houston, 429 Philip Guthrie Hoffman Hall, Houston, TX 77204-3011, USA. E-mail: fcantu10@uh.edu

[1]See *La Página*, February 16, 2014 and *Contrapunto*, February 21, 2014.

two largest parties in the country.[2] In Colombia, surveys conducted before the run-off presidential election were ambiguous and far from the final result, in which President Juan Manuel Santos was reelected with >50% of the vote.[3] Consequently, citizens often discredit the information provided by pollsters and consider them simply as another campaign tool for candidates.

The quest to assess the accuracy of polls has a long history, and the literature has grown rapidly in recent times with studies about the precision of preelection polls in the United States (Crespi, 1988; Lau, 1994; Mitofsky, 1998; Panagopoulos, 2009), the United Kingdom (Crewe, 1997, 2005; Curtice, 1997; Jowell, Hedges, Lynn, Farrant, & Heath, 1993), Ireland (McElroy & Marsh, 2003), France (Arzheimer & Evans, 2014; Durand, 2008; Durand, Blais, & Larochelle, 2004), Portugal (Magalhães, 2005), Germany (Schafer & Schneider, 2005), Australia (Jackman, 2005), Italy (Callegaro & Gasperoni, 2008), and New Zealand (Wright, Farrar, & Russell, 2014), to mention a few. Yet, for the most part, the literature has relied on developing a single statistic to evaluate polls.

Concerned with the peculiarities of polling in this region—multiparty systems, skepticism over survey data, and an ongoing debate among pollsters who consistently blame survey respondents, but never their methodologies, for inaccurate estimates—we take an alternate route that overcomes certain limitations of previously used measures for multiparty settings. With the benefit of hindsight, we track "true" vote intentions in the electorate throughout the two most recent presidential campaigns in Mexico, assess their differences relative to published estimates of vote intentions for each candidate at each point in time, and compute a measure of the *systematic error* for each candidate on each pollster's estimate.

The methodology we propose builds on Jackman's (2005) as an approach to estimate "house effects" in preelectoral surveys. Our contribution, aside from minor enhancements to model specification, is to demonstrate the utility of unpacking the bias that each polling firm produces *for each candidate* and showing that, in multiparty settings, "house effects" can be candidate-specific, and not generically pollster-specific (as they are traditionally construed). This feature is easily lost in the zero-sum nature of bipartisan settings. Perhaps for this reason, previous efforts have focused primarily on bias on the estimation of the winning candidates. In addition, our measures of systematic error computed for each pollster's estimates on each candidate competing in an election constitute a first step in evaluating and documenting the overall performance of electoral polling in this country.

[2]See *Semanario Universidad*, January 28, 2014.
[3]See *El Universo*, June 6, 2014.

Mexico is an interesting case for exploring poll accuracy for a couple of reasons. On the one hand, electoral surveys are, despite a deeply ingrained tradition of mistrust and vilification, a popular and ever growing source of political information (Camp, 1996; Kuschick, 2002). On the other hand, the performance of the Mexican polling industry as a whole has been rather inconsistent in recent years, further contributing to the confusion as to how useful they are (Moreno, Aguilar, & Romero, 2011).

The rest of this article proceeds as follows: Section 2 provides a brief description of the current political system and the public opinion industry in contemporary Mexico; Section 3 discusses the advantages of our measurement over other standard tools; Section 4 describes the data and estimation procedures; Section 5 shows the results and discusses their implications; and, finally, Section 6 concludes.

Mexican Elections and the Use of Polls

Although presidential elections were regularly and consistently held for most part of the 20th century, it was not until after the 2000 ballot—when the *Partido Revolucionario Institucional* (PRI) was ousted from office—that lingering doubts over the democratic nature of the regime were fully dissipated. Since then, national, state, and local elections have often been closely fought.

Mexican presidents are elected under direct, compulsory (no sanction), adult, universal suffrage every six years without reelection and by plurality rule.[4] The current party system is stable and dominated by three main national political parties: the formerly hegemonic PRI, the center-right *Partido Acción Nacional* (PAN), and the center-left *Partido de la Revolución Democrática* (PRD). Political parties are publicly funded, and there is a 3% national threshold for representation that means smaller parties often stand for one election and disappear for the next one.[5] Mexican electoral law regulates the equitable access to media outlets for all political parties and candidates competing in an election and ultimately bans the publication of all electoral polls three days before the election, supposedly with the intention of providing voters with a time for independent reflection.

[4]Mexico is federal system with a bicameral legislature. The lower chamber (Cámara de Diputados) is elected under a mixed-member majoritarian system for three-year terms (300 single seat districts elected by plurality and 200-list seats determined by nationwide PR) while the 128-member upper chamber (Cámara de Senadores) is composed of three members elected for each state (32) and another 32 more elected also by nationwide proportionality, all senators stand for six-year terms.

[5]The electoral authority, Instituto Federal Electoral (IFE), currently called Instituto Nacional Electoral (INE), determines the amount of public funding available for the parties each year according to a fixed formula. It also reviews the parties' expenditures and sanctions violations to spending caps, which change according to the office being elected. There were four small parties that met the threshold in the 2012 elections.

Surveys have only recently become a staple of Mexican political life. Despite the existence of brief stints in the 1940s (Moreno & Sánchez-Castro, 2009) and the 1960s (Almond & Verba, 1963), it was not until the 1990s that preelectoral polls started to be publicly discussed during presidential campaigns. Similarly to other emerging democracies and "late comers" to the use of opinion polls (Ansu-Kyeremeh, 1999; Reichmann & Schweiger, 2009), the institutionalization of surveys as a tool to measure electoral preferences has neither been easy nor free of criticism. Because of Mexico's authoritarian past, scholars and practitioners alike have cast serious doubts—not only about actual poll estimates but also, and perhaps even more pervasively, about how to interpret them—as suspicions arose that polls were but another instrument to galvanize support for then-ruling authoritarian PRI by hiding the erosion of its traditional base or by concealing electoral fraud (Basáñez, 1995).

Although the political conditions that gave rise to these concerns disappeared when Mexico became a democracy, the polling industry has failed to completely assimilate this reality, at least when explaining its performance. Instead of seriously reconsidering their original choices and assumptions and gauging their reliability and track record in light of the new political reality, an important number of Mexican pollsters have continued to focus on survey respondents as the source of "error" and "bias" in surveys instead of questioning the surveys themselves or their methodologies.

We argue that rather than obsessing about pollsters constantly overestimating PRI candidates or underestimating candidates of a given party regardless of the election, the more meaningful questions in need of an answer are: How systematic are these biases, for which candidates, and by which pollsters? No serious assessment of the performance of the Mexican polling industry can start off on the right foot without answering these questions.

In sum, the concerns about preelectoral poll performance that once centered on their contributions to the survival of an authoritarian regime have today become a question about the quality—and methodological soundness—of published surveys in democratic Mexico. Rather than having a single-number statistic for *all* estimates produced by a polling firm, we find it more informative to know whether a polling firm estimates none, one, two, or all three parties with bias, and the *magnitude* and *direction* of such bias.

Why a Different Measure?

To date, the majority of studies seeking to assess poll accuracy have relied on one of Mosteller's (1949) eight methods to measure errors in election forecasts. The most popularly used have been a measure of accuracy for the first place relative to the election outcome (Mosteller 1), the average of the absolute differences between each party's forecast and its actual performance on

Election Day (Mosteller 3), or the precision of the estimated advantage that the first place holds over the second place (Mosteller 5).

Seeking to enhance poll precision evaluations, Martin, Traugott, and Kennedy (2005) proposed A, a new measure that focuses on the accuracy of estimates for *one party* relative to another, and essentially indicates the direction of the bias.[6] But despite its many advantages, quantifying bias for one party relative to another using A only makes sense if the other party is estimated without bias (Arzheimer & Evans, 2014). As a result, Arzheimer and Evans (2014) propose B and Bw, alternative measures that account for the possibility of measurement error.

All these methods seek to compute a *single statistic* to summarize the accuracy of polling estimates. This is, undoubtedly, a valid pursuit when a quick glance at overall poll precision is all that is required. But as in all aggregation processes, some (useful) information is necessarily lost. A single statistic, for example, may be confounding when all biases do not affect estimates in the same direction.[7] Similarly, nothing dictates that all candidates should be equally overestimated or underestimated; as the number of candidates increases so do the possibilities and degrees of over- and underestimations. For this reason, in multiparty settings, we could find it useful to determine which candidates were measured with error, how much so, and by whom. Therein lies our interest for the Mexican case.

The method that we use here to evaluate poll accuracy has an additional advantage, which requires explaining it in greater detail. Mosteller-type measures of survey bias typically use survey estimates that are "sufficiently close" to the election, and compare them to electoral outcomes. Hence, they rely only on a single estimate for each election to assess poll accuracy.[8] But a survey is a sampling exercise regardless of the time in which it is conducted.

That is, by exclusively relying on point estimates, we would need to assume that there is no uncertainty in these estimates. That would only be possible if we had a census instead of a population sample as we do in a survey. But even if we consider intervals around the point estimates constructed using the margin of error for each of them, we would need to assume that each estimate has a confidence level of 100%, which is impossible in a survey because it comes from a sample. So it is perfectly plausible that a methodologically sound survey produced an unreasonable estimate by mere

[6]In addition, they sketch A as an alternative measure of bias on estimates of a third party, relative to the two main parties, in multiparty settings.

[7]That is certainly not the case in bipartisan systems as the overestimation of one candidate necessarily implies the underestimation of the other, because vote intention is reported as proportions of voters. But it would need serious reconsideration in multiparty settings.

[8]It is true that we could make these estimates "dynamic" by comparing point estimates across elections. But our interest in this article lies in using all available information *within an election* to compute an estimate of survey error per candidate per firm. This has the added benefit of providing more accurate estimates on each election that can make a dynamic exercise even more precise.

chance. As a matter of fact, with a confidence level of 95% that is typical in surveys, we would expect that this would happen five times out of a hundred. Thus, if we only look at the last estimates in the campaign for each polling firm, we could be erroneously concluding there is bias when indeed the estimate we look at happened to be off by mere chance. This is why we need to borrow (statistical) strength from using the full series of estimates of vote intentions generated by a polling house over the course of a campaign for each one of the competing candidates. Asymptotically, these types of random errors would cancel out and converge to the "true" estimate of bias for each polling firm for each candidate.

Data and Methods

Unlike existing methods, our proposed methodology uses the official electoral returns as an anchor to track "true" vote intentions throughout the campaign, quantify differences with published estimates at each point in time, and compute a measure of systematic error for each candidate on each pollster's estimates. We do not focus on whether a single estimate on a given date for an individual candidate was off, but rather on whether a polling firm biases its estimates systematically in a particular direction. Our measure can easily detect whether "outlier" estimates were due to random error, and, in those cases, it can detect and compute the magnitude of the "house effects," as well as the uncertainty surrounding our calculus. The analysis below is based on a Kalman filter model (Kalman, 1960), which tracks a moving target—the share of vote intentions for each candidate—over the course of the campaign, extracting all available information from all published polls throughout the campaign and effectively separating noise from signal.

To distinguish the potential bias in the estimations presented by every polling firm, we conceptualize the reported proportion of voters supporting each of the candidates as the combination of two factors: the candidate's true support level among voters and the house effect of the polling firm. The first element considers the number of voters that would cast a ballot for each candidate at the time of data collection, and its potential variation is due to the population sampling error. The second element measures the systematic bias of the polling firm to overestimate or underestimate the true support for each candidate. This effect arises when the estimations for a candidate are consistently more or less favorable across the surveys of a given polling house. The separation of these two elements allows us to differentiate the systematic effect from the random error in every poll.

In our analysis, we make three basic assumptions. First, the electoral results are the exact measurement of each candidate's support on election day. That is, we assume that the proportion of votes for each candidate is a valid

and reliable way to operationalize the true support estimates at the moment that the election was held. Second, true support on any given day is a backward-looking predictor of true support on the previous day. Finally, we consider that any daily variations of true support for the candidates should be marginal. In other words, we are not expecting support for a candidate to change significantly from one day to the next.[9] Based on these assumptions, we compute both the true support and the house effects for estimates on voter support for *each* of the three most popular candidates by *each* polling firm during the 2006 and 2012 presidential campaigns in Mexico.[10]

As an illustration of the methodology, consider the estimation of the true support for PRI's Enrique Peña Nieto in the election of July 1, 2012, whose proportion of valid votes was 0.39. We estimate his true support on June 30 as a random draw from a normal distribution with mean 0.39 and a standard deviation of 0.02.[11] Thereafter, we calculate the true support for Peña Nieto on June 29 as another random draw from a normal distribution centered on the estimation obtained on the election day's eve, and so on. The estimates of every parameter are based on 500,000 iterations for three Markov Chains Monte Carlo methods. The estimation of the house effects comes from computing the difference between reported support for a candidate on the poll and the random draw for the true support at the day in which the survey was published.

The data analyzed in this article includes all *published* preelection polls during the past five months of the presidential campaign; that is, from February until June of 2006 and 2012.[12] We use this period for illustration and comparability purposes, despite the different dates for the start of presidential campaigns in 2006 and 2012—January 19, 2006 and March 30, 2012—because the first published poll in 2006 appeared on February 11, almost a month after the campaign had started. Our analysis is constrained to the last two presidential elections given the relatively lower number of surveys for

[9]The Appendix specifies the model we used for the estimations. We also provide in the Supplementary Information additional sensitivity tests on the assumptions of our priors to measure polling house effects and daily variation of the true support.

[10]We estimate the model using a Markov Chain Monte Carlo (MCMC) Bayesian estimation with Gibbs sampling for three chains and 500,000 iterations. To build our priors, we use the proportion of votes for each candidate as the true support level on the day of the election. From that, we estimate the true support for each of the previous days before the election using a reverse random-walk, in which the estimation of day t depends on what it has been estimated on day $t+1$. We conceptualize the prior distributions of the house effects as normal distributions with mean 0 and a standard deviation of 0.05. These distributions, or our initial beliefs that no polling firm has a particular bias for any of the candidates, are updated with the analysis of the data. See Beck (1989), Jackman (2005), and Linzer (2013) for related approaches.

[11]As the supplementary material shows, the results hold when using different assumptions for the variance on the daily true support.

[12]The results below exclude the estimations for those candidates whose electoral support was below 5%. That was the case for Patricia Mercado and Roberto Campa in 2006, as well as Gabriel Quadri in 2012. Our model cannot distinguish any house effect for these low levels of reported electoral support. The Appendix contains the information for their estimated values, which are nonconvergent after 1 million iterations. Excluding the estimations of these candidates does not affect the estimations for the front-runners.

other electoral races in the country.[13] We rely on data entrusted by law to the Federal Electoral Institute (IFE) once published and who, in turn, places it in an official public repository.[14] These polls were validated with those of the data repository at the Mexican Association of Market Intelligence and Opinion Agencies (AMAI).[15] In total, there were 16 pollsters who published 58 preelection surveys in 2006, and only 13 pollsters who published 183 preelection polls in 2012.[16]

Results

With our model, we are able to produce estimates that had been previously unavailable for the 2006 and 2012 presidential elections in Mexico. On the one hand, we are able to present a clearer picture of "true" vote intention dynamics for each candidate throughout the campaign period.[17] On the other hand, we calculate the bias on each polling firm's estimates for each candidate. From these results, we conclude that the 2006 election was even more competitive than previously assumed, and that house effects on a number of candidates were prevalent throughout both elections.

The 2006 presidential election was the closest and most contested in contemporary history (Lawson, 2007). Felipe Calderón ran as the candidate of the incumbent center-right PAN, and Andrés Manuel López Obrador as the candidate of the left-leaning PRD, with both starting the race as frontrunners. Additionally, it was the first time in recent history that the PRI candidate Roberto Madrazo ran throughout the campaign in a distant third place and finished more than 12 points below the two main contenders.

These dynamics were unprecedented and made many analysts and pollsters uneasy, as they did not know how to interpret a possible and second non-PRI victory in a context where the PRI, for the first time in >70 years, was not even a relevant contestant. The race ended with both Calderón and López

[13]For example, between February and June of 2000, there were only 30 surveys published by 11 different pollsters, which significantly increases the estimated variance for our estimates. See Kuschick (2002).

[14]Available at http://www.ife.org.mx.

[15]Available at http://www.opinamexico.org.

[16]For the 2012 election, 57% of the polls in the analysis come from the Milenio (GEA-ISA) daily tracking from. To make sure that the estimations for this election are not biased by the inclusion of this pollster, the Appendix shows the results for the analysis when leaving out of the sample the information for the 105 of its surveys. The estimations of the house effects for the other pollsters do not change with the inclusion of this pollster.

[17]There were a number of journalistic efforts to produce (unweighted) polls of polls, which, failing to account for the possibility of asymmetrical measurement error for different candidates, were perhaps producing inaccurate estimates of vote intention dynamics. An exception to this approach is estimates for the 2012 election—akin to our own—constructed by Diego Valle-Jones and published on his blog, available at http://blog.diegovalle.net.

Obrador separated by a minimal difference (0.56% of the vote) and with many weeks of uncertainty as to who the actual winner of the election had been.

Figure 1a shows our estimates of the "true" share of vote intention throughout the campaigns for all three main candidates. The surrounding shadowed areas constitute the degree of uncertainty of our estimates. Our point estimates for 2006 suggest that López Obrador did start the campaign with a clear advantage over Calderón, which he lost at some point in early April, never to recover. But once we account for the uncertainty of these estimates, we see that the campaign began "too close to call" and remained as such throughout the race, as shown in the overlap of the shadowed areas. Moreover, the tight race between López Obrador and Calderón kept out of sight the systematic overestimation of Roberto Madrazo. Despite being third place for most of the electoral campaign, preelectoral surveys consistently provided higher estimates for the PRI's candidate, causing many analysts and the public opinion to be truly surprised by his then unexpected poor performance in the final results.

But how did pollsters fare at capturing these dynamics throughout the campaign? Figure 2a presents the distribution of the estimated systematic error—the "house effect"—on each polling firm's estimates. Our model initially assumes that the systematic error for each pollster follows a normal distribution with mean 0 and standard deviation of 0.05, and this assumption is updated after estimating the model. Hence, those distributions with <95% on either the left or right side of 0 are considered to have accurately estimated vote intention for a particular candidate throughout the campaign. The values for the 0.025, 0.500, and 0.975 quantiles for each distribution are in the Appendix.

On Figure 2a, we see that many pollsters estimated López Obrador rather accurately during the campaign. Thus, despite López Obrador's claims in 2006 of pollster bias against him, he was the best-estimated candidate in the campaign. Yet, the same was not true for Calderón: four pollsters—Consulta Mitofsky, Demotecnia, Indemerc and Parametria—who published a third of the surveys during the campaign consistently underestimated his share of vote intention. Additionally, the average overestimation for the PRI's candidate across pollsters ranges from three to six points above his true support level. Finally, we see that nine pollsters, responsible for more than a half of the preelectoral surveys during the campaign, consistently estimated all three candidates within acceptable ranges of accuracy throughout the campaign.

The 2012 election was not nearly as competitive as the 2006 race. This time around, the campaign began with PRI's Enrique Peña Nieto having a sizable advantage in vote intention. From this perspective, and compared with 2006, it was all smooth sailing for the frontrunner, who enjoyed a sizeable lead

Figure 1

Estimated vote intention per candidate throughout the 2006 and 2012 Mexican presidential election campaigns. Lines correspond to the estimated "true" vote intention for each candidate and the respective shaded areas represent the 95% credible interval of its posterior distribution. Pollster-specific dots represent polling firm estimates per candidate at a particular point in the campaign. The difference between the estimated line and the polling firm estimate corresponds to estimated polling house bias at a given point in the campaign.

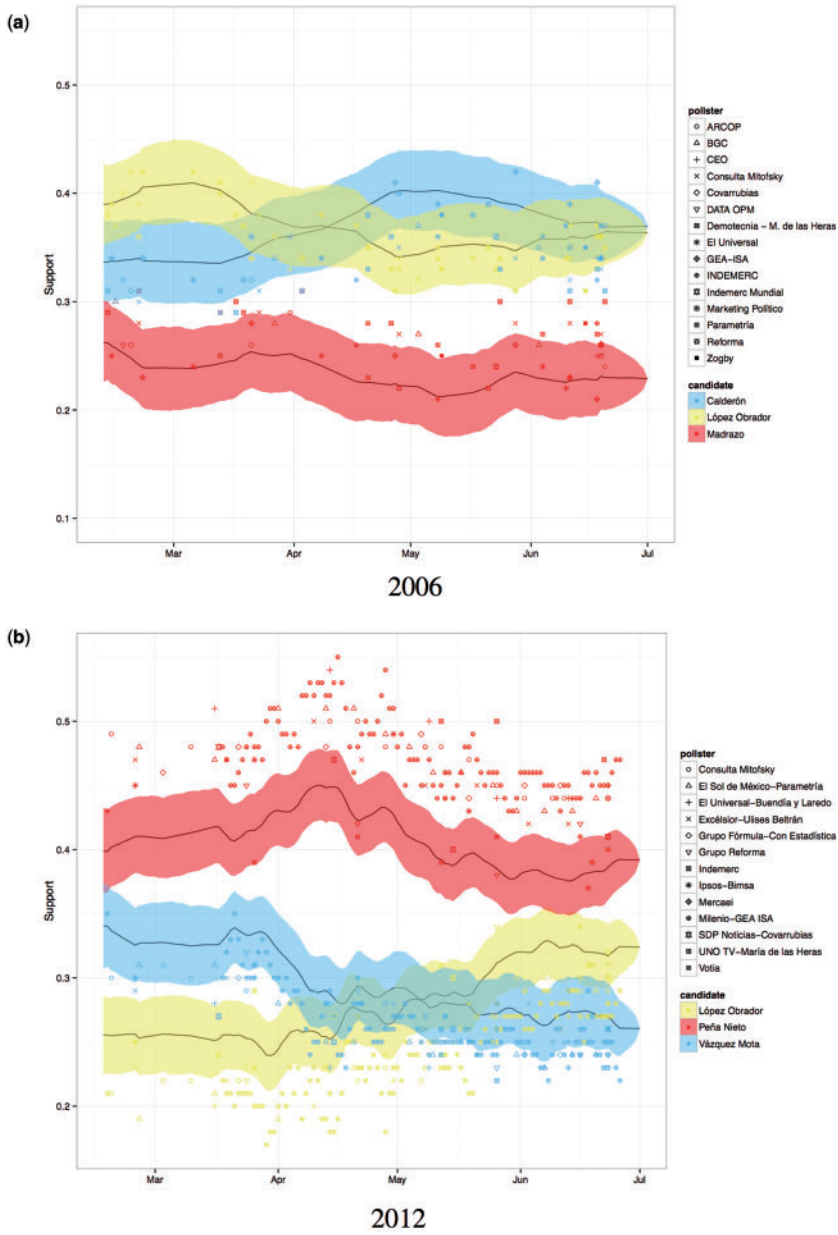
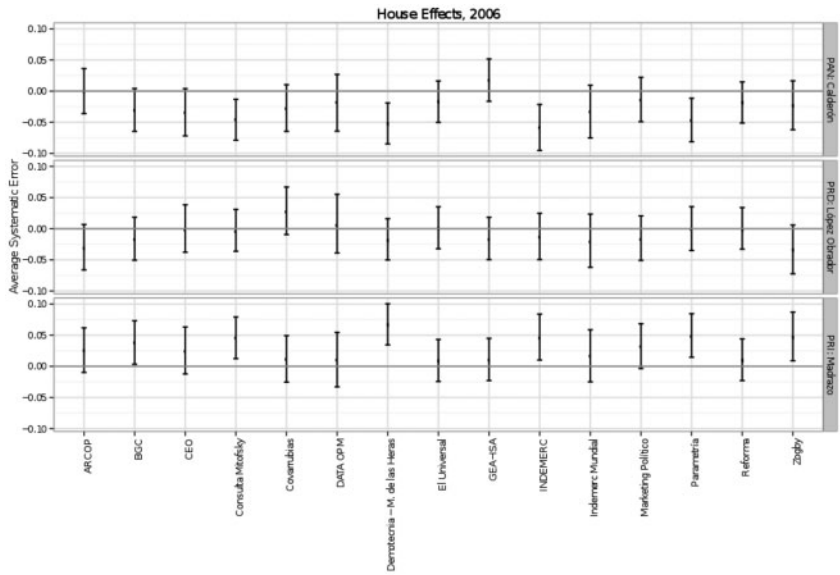
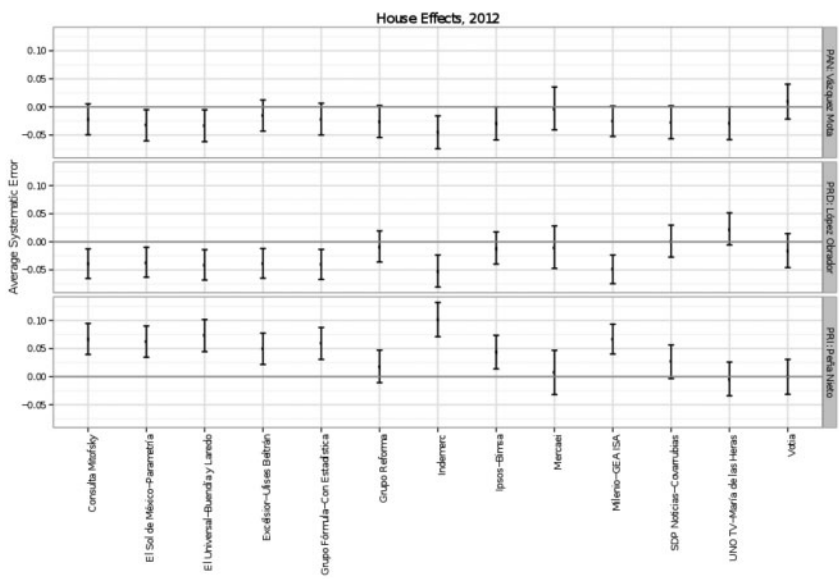


Figure 2

Estimates of systematic error per candidate per pollster for the 2006 and 2012 Mexican presidential elections. Distributions appearing on the positive (right) side of the graph imply that a pollster tended to overestimate a candidate throughout the campaign, and distributions on the negative (left) side imply underestimation.



(a) 2006



(b) 2012

over the second- and third-place candidates. Not surprisingly, Peña Nieto won the election with 38.21% of the vote, followed by PRD's Andrés Manuel López Obrador (31.6%), and PAN's Josefina Vázquez Mota (25.4%).[18]

Figure 1b shows these campaign dynamics. Although vote intention for Peña Nieto remained high throughout the campaign but diminished by the end of it, it was still high enough to give him a comfortable advantage. Vázquez Mota started in a close second but kept losing ground as the race progressed until she ended in a distant third place. López Obrador, who started third, gained traction by mid-April and continued to garner vote intention but only enough to make him finish second. This figure shows that the estimations of the pollsters were sensitive to different events during the campaign. For example, the largest raise for the declared support of López Obrador was in mid-May, after the student protests against Peña Nieto's campaign (Moreno & Dominguez, 2015). Similarly, the deepest decline in the declared vote for Vázquez Mota appeared during the first half of April, before the candidate changed her campaign team and strategy.[19]

Just as in the analysis for the 2006 election, Figure 2b reveals definite house effects for the 2012 presidential election. With the exception of five pollsters—Reforma, Mercaei, Covarrubias, María de las Heras, and Votia—all other polling firms overestimated vote intention for Peña Nieto. As expected, underestimation for the other candidates seems to have been more prevalent during the campaign. While seven pollsters—BGC, Consulta Mitofsky, Buendía y Laredo, GEA-ISA, Indemerc, Parametría, and Con Estadística, responsible for >80% of the preelectoral surveys—underestimated vote intent for López Obrador, three pollsters—Buendía y Laredo, Indemerc, and Parametría, publishing a combined total of 24 surveys—underestimated the declared support for Vázquez Mota.

In sum, there seemed to be a larger degree of bias in 2012 relative to 2006, which is surprising given that it was a much less complicated campaign to track and that there were many more polls conducted. Nevertheless, there is consistency on the estimated "house effects" between elections. Table 2 in the Appendix shows that some firms regularly overestimate voter support for the PRI's candidates. However, this observation cannot be generalized to the whole industry, especially because five pollsters estimated vote intention accurately for all candidates throughout the campaign.

[18]Gabriel Quadri from the New Alliance Party (PANAL) got 2.3% of the vote. We exclude his estimations from the summary of the results because his electoral support is lower than the assumed margin of error from the pollsters and their Monte Carlo Markov Chains did not converge even after a million iterations.
[19]See *CNN México*, April 9, 2012.

Table 1
Survey error statistics for 2006 and 2012 Mexican Presidential Elections

	Mosteller (1949)				Martin et al. (2005)		Arzheimer et al. (2014)	
Year	1	3	5	$\overline{A}'_{PRI}$	$\overline{A}'_{PAN}$	$\overline{A}'_{PRD}$	$\overline{B}'$	$\overline{B}'_{w}$
2006	2.37	2.73	3.88	0.181	0.095	0.056	0.129	0.119
2012	4.05	3.07	1.76	0.128	0.103	0.053	0.140	0.145

Comparison with Other Methodologies

We have argued that having candidate-specific estimates of systematic house effects throughout the campaign for each polling firm provides a rich perspective of the performance of the polling industry in multiparty settings in ways that the commonly used single-statistics cannot. To illustrate this difference, we computed the most common single statistics of polling accuracy for the same Mexican elections and present them in Table 1. Our computations use the latest estimates published by each polling house for the three main candidates during each campaign (typically in mid-June), and the actual election results.

We calculated Mosteller's (1949) three most popular measures of polling accuracy: *Mosteller 1*, which captures the average of the differences between each polling firm's forecast for the winning candidate and the election result; *Mosteller 3*, which focuses on the average of the absolute differences between the vote share predicted by each polling firm for each candidate and the election results; and *Mosteller 5*, which is the average of the differences between the forecast and the election result for the winner and the second place as presented by each polling firm. The interpretation of these statistics is straightforward: Positive (negative) numbers indicate overestimation (underestimation) of the winning party (*Mosteller 1*); positive numbers indicate the magnitude of overall inaccuracy in forecasts (*Mosteller 3*); or numbers different from zero indicate increased imprecision in the estimation of the winner and the second place (*Mosteller 5*).

In addition, we computed Martin, Traugott, and Kennedy's (2005) A statistic as modified by Arzheimer and Evans (2014) for multiparty settings, thus providing an estimate of bias for party i relative to the remaining two parties, dubbed A'_i. Its interpretation is also straightforward: Positive values indicate bias in favor of party i, while negative values indicate bias against this party. Finally, we estimate Arzheimer and Evans' (2014) B statistic, which is an unweighted average of the absolute values of the estimated A'_i, as well as Bw, which extends B by weighting each candidate's estimates by their proportion of the total votes. The interpretation of these statistics is similar to the previous ones, in that positive (negative) values express the average factor by which candidates are overestimated (underestimated).

The aggregate nature of these statistics produces a useful picture of the behavior of polls in any given election. For instance, from statistics estimated for the 2006 presidential campaign, we can infer that on average: The winning PAN candidate was underestimated (Mosteller 1, $A_{PAN}\hat{}$); the (second place) PRD candidate was underestimated ($A_{PRD}\hat{}$); PAN and PRD candidates were not very well estimated (Mosteller 5); the PRI candidate was overestimated ($A_{PRI}\hat{}$), which implies that all candidates were inaccurately forecasted on average (Mosteller 3); and all surveys had a degree of bias (B, B_w).

Similarly, from the statistics for the 2012 presidential campaign we can infer that on average: The winning PRI candidate was overestimated (Mosteller 1, $A_{PRI}\hat{}$); both PRD and PAN candidates were underestimated ($A_{PRD}\hat{}$, $A_{PAN}\hat{}$); PRI and PRD candidates were not very well estimated (Mosteller 5); all candidates were inaccurately forecasted (Mosteller 3); and all surveys had a degree of bias (B, B_w).

Yet, despite their usefulness to detect aggregate shortcomings in forecasting electoral outcomes, there are some important limitations in these statistics that are worth unpacking. First, these statistics rely on a single observation—typically the poll closest to election day—to estimate bias. With this information, it is hard to distinguish measurement error from sampling error for each individual poll that goes into the estimated statistic. It could be entirely possible that we surmise as bias an estimate that is off by mere random error, which we could not distinguish from specific sampling procedures of that single survey. Even when there have been efforts to widen the time frame to better capture vote intent dynamics (Durand, 2008), the problem of having an accurate estimate of vote intention at each point in time for each candidate, which effectively separates sampling error from measurement error, remains unsolved.

Second, aggregating all information into a single statistic lumps together biased and accurate estimates. That is, we can characterize performance of surveys in an election as a whole, but we cannot differentiate the individual firms that consistently estimate a candidate accurately from those that do not. Even when the literature has recently tended to provide individual calculations for each one of the firms producing survey estimates in a campaign (Arzheimer & Evans, 2014; Durand, 2008), they remain heavily dependent on election day results and not on estimates of vote intention at different points in the campaign.

The method we rely on addresses these concerns simultaneously, as it makes a more efficient use of all available information, allowing the estimation of true vote intention and survey bias throughout the entire campaign. In addition, having multiple measures of poll accuracy for each candidate competing in an election provides a fuller picture, and the ability to pinpoint direction and magnitude of survey bias.

In the case of Mexico, we have clear evidence that all "house effects"—when we look at each candidate's estimates separately—are not created equal. In fact, we see that in multiparty elections, it is entirely possible that some pollsters systematically overestimate certain candidates, underestimate others, and also estimate some with no bias. Evidently, the best performing pollster should be that who estimates all candidates accurately. But knowing which pollsters estimate which candidates inaccurately is also useful information for consumers to discount, and for pollsters to correct. We are able to do this because we can produce measures of systematic error throughout the campaigns. Evidently, this underscores the utility of unpacking single-statistic evaluations of poll accuracy in multiparty settings.

Beyond the specific dynamics of "house effects" on each presidential campaign, we can also learn something from the collective performance of these biases. In the Mexican case, we see that they seem to have moved in the same direction. That is, there were no mixed bags in the collective dynamics of systematic error. Among pollsters who got it wrong, all tended to overestimate or underestimate candidates in tandem, a finding that is remarkable given the divergence in methodologies and sampling procedures across polling firms.

Finally, if we look closer, there are other discernible patterns across our results. In the last two Mexican presidential races, there were no frontrunner or underdog effects, as the frontrunner was underestimated in 2006 and overestimated in 2012 and the trailing candidate was overestimated in 2006 and underestimated in 2012 by the majority of pollsters. Instead, we see a much clearer "party effect," such that PAN candidates seem to be underestimated by a majority of pollsters, while PRI candidates are always overestimated. Again, this is true only within our limited time frame of only two elections and further analysis is warranted.[20]

These results may have a number of explanations, both methodological and logistical. A preliminary analysis suggests a mild correlation between refusal rates and increased systematic error, yet results are not robust enough to warrant further comment. We suspect that one possible reason may be related to logistical constraints when polls are fielded that hinder the quality of the survey estimates. Most pollsters in Mexico do not have a permanent staff of interviewers. Instead, they rely on various local providers to carry out their polls. This being the case, it would not be strange that more than one polling firm relies on the same local provider, which might overstretch human resources when demand is high, such as in electoral periods.

[20]Yet, the possibility of these investigations is limited by the lack of relevant information to carry out these analyses. Even when, by law, every pollster that publishes an electoral survey must submit its results and a methodological note, the lax interpretation of what constitutes a methodological note makes it almost impossible to know the sampling frame for all published surveys.

Furthermore, when many daily surveys must be carried out for multiple pollsters, it may become harder to follow the protocols established by each national polling firm to deal with survey respondents when not found. Instead of making repeated attempts to find the selected respondent, the local interviewer may choose to replace her with the first available respondent in a household. This may naturally bias the sample toward people that are more likely to be at home during the day, and could potentially explain why PAN candidates are underestimated while PRI candidates are overestimated by a number of polling firms. Unfortunately, polling firms do not typically have additional oversight in place to verify whether this is the case, and even if they do it is unlikely they would release this information.

In terms of looking for the best poll accuracy measure, we argue that, had we only estimated the traditional, single-statistic measures of polling accuracy on Table 1, the finesse of the campaign dynamics that our proposed method detects would have been completely lost. Having these dynamics spelled out with clarity can be particularly relevant in multiparty settings where the natural kinematics of electoral competition make accurate vote intention estimates an obvious necessity.

Conclusion

Single-statistic measures of polling accuracy have been used extensively in the literature as means to distill a single number that rapidly informs about bias in polls. But, when applied in multiparty settings, these statistics have two major shortcomings: (1) they may not clearly distinguish the direction and magnitude of these biases *for each candidate* that competes in an election; and (2) they fail to exploit all available information risking confusing true random error naturally embedded in survey design with polling bias. We therefore advocate unpacking single statistics into estimates of polling error on each candidate by each polling firm with a method that makes full use of all available polling information and distinguishing between random and systematic error.

The utility of this strategy is illustrated here on the 2006 and 2012 presidential election campaigns in Mexico. We not only show that it is entirely possible that bias in a polling firm's estimates may not be unidirectional. Rather, it seems common—at least in our Mexican example—that the same firm may simultaneously overestimate some candidates, underestimate some others, and accurately estimate some more. We also show that it is possible that the direction of bias in estimates for a specific candidate may be shared by a number of polling firms in multiparty settings. These dynamics emphasize the importance of capturing the performance of each polling firm relative to each candidate. After all, this knowledge may help pollsters better correct their methodology or logistics, while simultaneously allowing voters to discount

biases when incorporating this information to cast their (sincere or strategic) votes.

In contrast to most studies of poll accuracy for the 2012 Mexican presidential election (Zuckermann, 2012a, 2012b), our estimation overcomes some of the obvious limitations of traditional methods. Instead of only considering the last polls (or those surveys closest to election day) for every polling firm, our model takes into account all surveys reported during the five months of the presidential campaign. A larger sample of surveys per polling firm helps us to distinguish the systematic effect from random error more precisely. Moreover, the operationalization of the house effects that we propose ignores noisy characteristics related to the quality of the polling firm, thus increasing the face validity of the model.[21] In sum, this methodology presents a way to clean out the effect of the polling firms from the true support of the candidates, allowing us to observe variation for the candidates' support throughout the campaign beyond what the surveys report.

Supplementary Data

Supplementary Data are available at *IJPOR* online.

Acknowledgments

We thank René Bautista, Jerónimo Cortina, Gary Jacobson, Seth Hill, Michael McDonald, Devesh Tiwari, Duncan Wood, and participants in UC San Diego's Center for US-Mexican Studies (USMEX) Wednesday Seminars and the 74th Annual Meeting of the Midwest Political Science Association for comments on earlier versions of this article. Morales conducted research for this project while he was a Fellow at USMEX and Hoyo a Visiting Scholar at the Center for Comparative Immigration Studies (CCIS) at UCSD.

[21]Besides the obvious choices of sampling design and weighting procedures, preelection polls (and their related forecasts), may be trampled by actual implementation issues such as the mode of the interview, the day of the week, and time of the day in which it happens, the age, gender, and race of the interviewer, question wording and question ordering McDermott and Frankovic (2003), Moreno et al. (2011), and Romero (2012). None of these will be considered in this article given the nature of the data that we use.

Appendix

Model Specification

Consider a set of different electoral polls, $k = 1, \ldots, K$. Each poll k is produced by polling firms $j[k]$, and it reports the number of respondents supporting candidate i at day t. Let $t = 1, \ldots, T$ index the campaign days so that $t = 1$ is the first day of campaign and $t = T$ is the election day. On any given day of the campaign, the number of voters supporting candidate i given poll k is given by the state vector

$$y_{i[k]} \sim \text{Binomial}\left(\pi_{i,j[k],t[k]}, n_k\right)$$

where n_k is the sample size of poll k. The reported support for each of the candidates, $\pi_{i,\ j[k],t}$ is modeled as a function of the true support for candidate i in the election at time t, $\alpha_{i,t}$, and the house effect by the polling firm j for or against the candidate, $\delta_{i,j}$. Therefore, we produce a measurement equation

$$\pi_{i,j,t} = \text{logit}^{-1}\left(\alpha_{i,t} + \delta_{i,j}\right)$$

A value of $\delta_{i,j} > 0$ represents a systematic overestimation by polling firm j to the support of candidate i. On the other hand, $\delta_{i,j} < 0$ shows a systematic underestimation to the true support of candidate i by polling firm j. Our priors for $\delta_{i,j}$ are normally distributed with a mean 0 and a standard deviation of σ. For the case of $\alpha_{i,t}$, we anchor the values at day T to the proportion of votes each of the candidates obtained in the election. Meanwhile for the $t = 1, \ldots, T$ 1 days of the electoral campaign, we model $\alpha_{i,t}$ as a normal distribution with mean $\alpha_{i,t+1}$ and a precision $\tau = s^2_i$. Therefore, our transition equation is modeled as a reverse random walk, specified as

$$\alpha_{i,t} | \alpha_{i,t+1} \sim N\left(\alpha_{i,t+1}, s^2\right)$$

We assign a value for τ given a uniform distribution over the interval [0, 0.05], while σ has a value of 0.05. For the estimation of the house effects, $\delta_{i,j}$, our priors for each polling firm and candidate are modeled as $\delta_{i,j} \sim N(0,0.05)$. The results show the posterior distributions of $\alpha_{i,t}$ and $\delta_{i,j}$ after 500,000 iterations and three chains. We use Geweke diagnostics to test for convergence of the chains.

References

Almond, G., & Verba, S. (1963). *The civic culture: Political attitudes and democracy in five nations*. Princeton, NJ: Princeton University Press.

Ansu-Kyeremeh, K. (1999). The challenges of surveying public opinion in an emerging democracy. *International Journal of Public Opinion Research, 11*(1), 59–74. doi: 10.1093/ijpor/11.1.59.

Arzheimer, K., & Evans, J. (2014). A new multinomial accuracy measure for polling bias. *Political Analysis, 22*(1), 31–44. doi: 10.1093/pan/mpt012.

Basáñez, M. (1995). Problems in interpreting electoral polls in authoritarian countries: Lessons from the 1994 Mexican election. *International Social Science Journal, 146*(4), 663–650.

Beck, N. (1989). Estimating dynamic models using Kalman filtering. *Political Analysis, 1*(1), 121–156. doi: 10.1093/pan/1.1.121.

Callegaro, M., & Gasperoni, G. (2008). Accuracy of pre-election polls for the 2006 Italian parliamentary election: Too close to call. *International Journal of Public Opinion Research, 20*(2), 148–170. doi: 10.1093/ijpor/edn015.

Camp, R. A. (Ed.). (1996). *Polling for democracy: Public opinion and political liberalization in Mexico*. Wilmington, DE: Scholarly Resources.

Crespi, I. (1988). *Pre-election polling. Sources of accuracy and error*. New York: Russell Sage.

Crewe, I. (1997). The opinion polls: confidence restored? *Parliamentary Affairs, 50*(4), 569–587. doi: 10.1093/oxfordjournals.pa.a028756.

Crewe, I. (2005). The opinion polls: The election they got (almost) right. *Parliamentary Affairs, 58*(4), 684–698. doi: 10.1093/pa/gsi074.

Curtice, J. K. (1997). So how well did they do? The opinion polls in the 1997 election. *Journal of the Market Research Society, 39*(3), 449–461.

Durand, C. (2008). The polls of the 2007 French presidential campaign: Were lessons learned from the 2002 catastrophe? *International Journal of Public Opinion Research, 20*(3), 275–298. doi: 10.1093/ijpor/edn029.

Durand, C., Blais, A., & Larochelle, M. (2004). The polls in the 2002 French presidential election: An autopsy. *Public Opinion Quarterly, 68*(4), 602–622. doi: 10.1093/poq/nfh042.

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

Jowell, R., Hedges, B., Lynn, P., Farrant, G., & Heath, A. (1993). The 1992 British election: The failure of the polls. *Public Opinion Quarterly, 57*(2), 238–263. doi: 10.1086/269369.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (Series D), 82*, 35–45. doi: 10.1115/1.3662552.

Kuschick, M. (2002). México: Elecciones y el uso de las encuestas preelectorales. [Mexico: Elections and the use of preelectoral surveys]*Revista Mexicana de Ciencias Polticas y Sociales, XLV* 103–127. [Spanish].

Lau, R. R. (1994). An analysis of the 'trial heat' polls during the 1992 presidential election. *Public Opinion Quarterly, 62*(1), 2–20. doi: 10.1086/269405.

Lawson, C. (2007). How did we get here? Mexican democracy after the 2006 elections. *Political Science and Politics, 40*(1), 45–48. doi: 10.1017/s1049096507070084.

Linzer, D. (2013). Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association, 108*(501), 124–134. doi: 10.1080/01621459.2012.737735.

Magalhães, P. C. (2005). Pre-election polls in Portugal: Accuracy, bias, and sources of error, 1991–2004. *International Journal of Public Opinion Research, 17*(4), 399–421. doi: 10.1093/ijpor/edh078.

Martin, E. A., Traugott, M. W., & Kennedy, C (2005). A review and proposal for a new measure of poll accuracy. *Public Opinion Quarterly, 69*(3), 342–369. doi: 10.1093/poq/nfi044.

McDermott, M. L., & Frankovic, K. A. (2003). Review: Horserace polling and survey method effects: An analysis of the 2000 campaign. *Public Opinion Quarterly, 67*(2), 244–264. doi: 10.1086/374574.

McElroy, G., & Marsh, M. (2003). Why the opinion polls got it wrong in 2002. In M. Gallagher & P. Mitchell (Eds.), *How Ireland voted in 2002* (pp. 159–176). London: Palgrave.

Mitofsky, W. J (1998). Was 1996 a worse year for polls than 1948? *Public Opinion Quarterly, 62*(2), 230–249. doi: 10.1086/297717.

Moreno, A., Aguilar, R., & Romero, V. (2011). La precisión de las encuestas electorales en México: Un análisis de las fuentes de error. [Electoral survey precision in Mexico: An analysis of sources of error] [Spanish]. *Revista Legislativa de Estudios Sociales y de Opinión Pública, 4*(8), 7–45.

Moreno, A., & Díaz-Domínguez, A. (2015). Effects of yosoy132 and social media in Mexico's 2012 presidential campaigns. In J. I. Domínguez, K. F. Greene, C. H. Lawson & A. Moreno (Eds.), *Mexico's Less Than Perfect Democracy: A Comparative Study of the 2012 Elections* (pp. 227–251). Baltimore, MD: John Hopkins University Press.

Moreno, A., & Sánchez-Castro, M. (2009). A lost decade? László Radványi and the origins of public opinion research in Mexico, 1941-1952. *International Journal of Public Opinion Research, 21*(1), 3–24. doi: 10.1093/ijpor/edp002.

Mosteller, F. (1949). Measuring the error. In F. Mosteller, H. Hyman, P. J. McCarthy, E. S. Marks & D. B. Truman (Eds.), *The Pre-election Polls of 1948: Report to the Committee on Analysis and Pre-Election Polls and Forecast* (pp. 54–80). New York: Social Science Research Council.

Panagopoulos, C. (2009). Polls and elections: Pre-election poll accuracy in the 2008 general elections. *Presidential Studies Quarterly, 39*(4), 896–907. doi: 10.1111/j.1741-5705.2009.03713.x.

Reichmann, W., & Schweiger, M. (2009). How Hayek managed to beat Lazarsfeld: The different perceptions of two sub-fields of social science. In C. Fleck, A. Hess & E. S. Lyon (Eds.), *Intellectuals and their publics: Perspectives from the social sciences* (pp. 243–260). Burlington, VT: Ashgate.

Romero, V. (2012). Notas para la evaluación de las encuestas preelectorales: las elecciones para gobernador de 2010 en México. [Notes to evaluate preelection surveys: Mexican gubernatorial elections in 2010] [Spanish]. *Política y Gobierno, XIX*(1), 101–124.

Schafer, L.-M., & Schneider, G. (2005). Die prognosegte von wahlbo rsen und meinungsumfragen zur bundestagswahl 2005. [The predictive power of electoral polls in the 2005 federal election] [German]. *Politische Vierteljahresschrift, 46*(4), 674–681. doi: 10.1007/s11615-005-0307-9..

Wright, M. J., Farrar, D. P., & Russell, D. F. (2014). Polling accuracy in a multiparty election. *International Journal of Public Opinion Research, 26*(1), 113–124. doi: 10.1093/ijpor/edt009.

Zuckermann, L. (2012a). ¿Quién es quién en las encuestas? [Who's who among pollsters?]. *Este País, 141*, 40–51. [Spanish].
Zuckermann, L. (2012b). ¿Quién es quién en las encuestas después de la elección de 2012? [Who's who among pollsters after the 2012 election?]. *Este País, 418*, 15–22.

News Sources

La Página, "FMLN aumenta su ventaja a 16 puntos para segunda vuelta, según CID-Gallup [FMLN increases its advantage by 16 points on the runoff, according to CID-Gallup]," February 16, 2014. Retrieved from http://www.lapagina.com.sv. [Spanish].
Contrapunto, "Ventaja de 20 puntos para el FMLN en encuesta UFG [20-point advantage for FMLN on UFG survey] [Spanish]," February 21, 2014. Retrieved from http:// www.contrapunto.com.sv.
Semanario Universidad, "Nueva encuesta del CIEP: Indecisión crece a pocos días de las elecciones [New CIEP survey: Indecision on the rise a few days before the election]," January 28, 2014. Retrieved from http://www.informa-tico.com. [Spanish].
El Universo, "Encuestas dispares para segunda vuelta electoral en Colombia [Dissimilar surveys in the Colombian runoff] [Spanish]," June 6, 2014. Retrieved from http://www.eluniverso.com/noticias.
CNN Mexico, "Vázquez Mota admite que los conflictos en el PAN le hicieron perder tiempo [Vázquez Mota admits that conflicts within PAN made her waste time] [Spanish]," April 9, 2012. Available at http://mexico.cnn.com/nacional.

Biographical Notes

Francisco Cantú is Assistant Professor in Political Science at the University of Houston. His research interests are comparative politics, political economy, and quantitative methods.

Verónica Hoyo has a PhD in Political Science from the University of California, San Diego. Her research is centered on the role of political parties, their strategies, and the nature of voter preferences in contemporary democratic politics. Her work has been published in *Electoral Studies*.

Marco A. Morales is Affiliated Researcher at *Instituto Tecnológico Autónomo de México* (ITAM). His current research focuses on refining the tools that have traditionally been used to measure and model economic voting, both at the individual and at the aggregate levels. His work has been published in *Electoral Studies*.