Public Opinion Quarterly Advance Access published April 14, 2015

Public Opinion Quarterly

THE POLLS–REVIEW
PREDICTING ELECTIONS: CONSIDERING TOOLS TO POOL THE POLLS

JOSH PASEK*

Abstract Surveys have long been critical tools for understanding elections and forecasting their results. As the number of election surveys has increased in prevalence, researchers, journalists, and standalone political bloggers have sought to learn from the wealth of information released. This paper explores three central strategies for pooling surveys and other information to better understand both the state of an election and what can be expected when voters head to the polls. Aggregation, predictive modeling, and hybrid models are assessed as ways of improving on the results of individual surveys. For each method, central questions, key choices, applications, and considerations for use are discussed. Trade-offs in choices between pooling strategies are considered, and the accuracies of each set of strategies for forecasting results in 2012 are compared. Although hybrid models have the potential to most accurately pool election information and make predictions, the simplicity of aggregations and the theory-testing capacity of predictive models can sometimes prove more valuable.

Every four years, American media engage in a curious ritual; over the course of a few months, they unleash survey after survey upon the American public to query choices in an upcoming presidential election. Broadcasters hope that opinion polls will help them predict election results and track the impacts of evolving electoral strategies. But surveys are a blunt instrument. Although they capture a snapshot of the attitudes of the public, typical election polls are not sufficiently precise to measure small day-to-day changes in the performance of contestants (Erikson and Wlezien 1999; Jackman 2005); they also

Josh Pasek is an assistant professor of communication studies and faculty associate in the Center for Political Studies, Institute for Social Research, at the University of Michigan, Ann Arbor, MI, USA. The author thanks Mike Traugott, Drew Linzer, and Mark Blumenthal for useful comments on the manuscript. Drew Linzer, Mark Blumenthal, Andreas Graefe, and Pollster's API provided data necessary for accuracy comparisons between methods. *Address correspondence to Josh Pasek, University of Michigan, 105 S. State Street, 5413 North Quad, Ann Arbor, MI 48109; USA; phone: (734) 764–6717; e-mail: jpasek@umich.edu.

doi:10.1093/poq/nfu060

reveal citizen preferences at the moment of the survey rather than the behaviors that might be expected on election day, limiting their forecasting potential (cf. Erikson, Panagopoulos, and Wlezien 2004; Traugott 2005; Moore 2008). Further, a number of sources of error—from sampling variance and question wording to interviewer characteristics and nonresponse—can produce misleading survey results (Walsh, Dolfin, and DiNardo 2009; Pasek and Krosnick 2010; Groves and Lyberg 2011).

Practitioners have long sought ways to improve upon the performance and predictive capabilities of election surveys. The pages of this journal have been employed both to document methodological improvements and to provide retrospectives when survey estimates diverged from official tallies (Crossley 1937; Katz 1941; Wilks et al. 1948; Gallup 1951; Mitofsky 1998). Methods for sampling and conducting election surveys have become considerably more rigorous since the straw polls of the 19th century and the quota sample surveys that dominated the field in the 1930s and 1940s (Converse 1987; Herbst 1993; Berinsky 2006; Hillygus 2011). Yet, numerous challenges remain. Polling firms make decisions about how and when people should be surveyed, which respondents should be classified as likely voters, and how sampled individuals should be weighted to represent the public (Voss, Gelman, and King 1995; Hillygus 2011). Hence, differences between survey results and election outcomes may be a function of sampling error or any of a litany of design decisions.

Combining data across polls and other data sources is one way to mitigate idiosyncratic errors in individual surveys (cf. Erikson and Wlezien 1999; Green, Gerber, and De Boef 1999; Jackman 2005; Hillygus 2011; Graefe et al. 2014b). Three basic approaches empower researchers and poll watchers to pool estimates when tracking elections or projecting election outcomes: aggregation, predictive modeling, and hybrid modeling. Aggregation enables researchers to use results from multiple surveys to produce a single estimate of the state of an election. This strategy is employed when news outlets report the "poll of polls" and by sites like RealClearPolitics. Electoral prediction models enable researchers using surveys and other data to forecast results given current dynamics; they provide a picture of what might be expected on election day by discounting known sources of error. Electoral predictions are primarily seen in the academic literature, most notably in a quadrennial pre-election issue of *PS: Political Science and Politics* (see Campbell 2004, 2008, 2012). Finally, hybrid approaches allow researchers to incorporate far more information into their estimates than would be possible with either aggregation or prediction models alone. These models combine multiple estimation or prediction approaches and are used in the estimates provided by the Huffington Post's Pollster and Nate Silver's FiveThirtyEight.

This paper outlines key distinctions between three approaches to combining survey data, highlights choices that practitioners must make, and stresses advantages and limitations of each estimation strategy. Aggregation, prediction, and hybrid approaches are explored, and conditions under which they might be expected to yield notably similar or different conclusions are discussed. The methods are considered with an eye toward their ability to capture current election dynamics and project election day results. Although the manuscript focuses on aggregating preferences at the state and national levels, translating results into electoral college projections would not alter any central considerations between the methods presented.

Beyond the approaches examined herein, additional strategies for predicting and tracking elections fall outside the scope of the current comparison. I do not address electronic betting markets, large panel surveys of elections, non-trial-heat measures of public opinion, or forecasting models that do not pool public opinion data.[1]

Aggregating Survey Data

Ideally, all survey data should accurately reflect population preferences; in practice, any individual survey is likely to fall short because of two basic types of error. Random variations, such as sampling error, ensure that any given survey estimate may differ from the population of interest by a few percentage points. Systematic sources of error, including coverage errors, method biases, and response biases, can cause all surveys of a particular type to misestimate a parameter of interest. For various reasons, averaging estimates across surveys can be a useful strategy for mitigating both types of error. Incorporating multiple data points limits random error by reducing the uncertainty of the estimates. If individual surveys vary in the forms of systematic error that are present, then aggregating across surveys will also tend to discount any single non-universal bias. For example, errors in interactive voice response surveys, which miss the cell-phone-only population, are unlikely to correspond with those of opt-in web surveys, which exclude offline individuals, when the two are averaged.

The strategy of aggregating data across polls to produce more accurate estimates has a long history in election analysis (Abrams 1970; Armstrong 2001);

1. Electronic markets were excluded for three reasons: a) their capacity has been discussed at length in earlier research (Bohm and Sonnegard 1999; Kou and Sobel 2004; Erikson and Wlezien 2008a; Rothschild 2010; Tziralis and Tatsiopoulos 2012), b) they are not centrally about survey data, and c) they may be subject to manipulation (Hansen, Schmidt, and Strobel 2004). In practice, however, manipulation appears to be rare and short-lived (Snowberg, Wolfers, and Zitzewitz 2012; Berg and Rietz 2014). Large panel surveys of elections also fall out of scope because they are principally used to understand elections post hoc, rather than as they are happening (Sides and Vavreck [2013] is a notable exception). Non-trial-heat survey measures such as citizen forecasting and presidential approval, which can be used to predict election outcomes, are also outside the scope of the current comparison for the sake of parsimony. Finally, many election forecasting models do not include (or do not pool) survey data (cf. Lewis-Beck 2005; Bélanger and Soroka 2012; Campbell 2012); these are not discussed in the current article, as the focus is solely on improving upon survey-based estimates.

principles behind this approach should be familiar to any lay statistician. Combining estimates from multiple samples is in many ways the equivalent of collecting a larger sample. Larger samples, in turn, suffer from less sampling error than smaller ones (though not necessarily less total error; see Bartels [1996]).[2] The mean for an aggregated estimate can be produced as the average of all estimates in a time period weighted by the precision with which each estimate was generated.[3] Aggregating also provides an advantage for survey consumers who are unaware of which estimates can be trusted. An aggregated estimate will typically be closer to the actual result than the average estimate that was aggregated (Mosteller 1948; Armstrong 2001).[4] But caution is merited, as a single unbiased, high-quality survey will frequently outperform an aggregation of biased, lower-quality results.

CHOICES IN AGGREGATION

Although aggregating polls is a relatively simple endeavor, aggregations can be produced using a variety of strategies. One technique involves simply taking the statistical mean of topline results for a set of surveys. Aggregators can improve on these estimates by considering the sample sizes and precisions of similarly conducted polls. All else being equal, large sample surveys will produce more precise estimates than small sample surveys. Hence, weighting methodologically similar results in proportion to the square root of their sample size should decrease the expected errors of the aggregated result. Smoothing algorithms can also be used to estimate how much of the change observed from survey to survey can be attributed to changing opinions rather than sampling error (Green, Gerber, and De Boef 1999).

2. As the total number of respondents gets larger, estimates become increasingly precise. Whereas two samples with 500 respondents might each have a margin of error of four percentage points, a single sample with 1,000 respondents yields a confidence interval of no more than 3.1 percentage points (Cochran 1953).

3. For surveys with similar methodologies, the precision of each estimate is inversely proportional to the square root of the size of the sample used for the estimate. Notably, given that the sampling error may not be the only source of error (and thus that error may not be inversely related to sample size across methods and houses), it may be equally or even more appropriate to simply average a series of estimates without weighting to account for the sample sizes of surveys from different houses. Three possible approaches would improve on a naïve averaging: 1) researchers could produce a precision-weighted average of surveys conducted using similar methods, 2) researchers could parameterize precision by assessing how results from each method have historically compared to final election tallies, and 3) precision could be estimated by the variation observed across surveys from a single firm. For most states, there is not enough data to produce meaningful unique estimates of these parameters.

4. This is true by definition from a mean squared error perspective. The smallest possible mean squared error for any set of estimates is equal to the squared error of the average estimate. From an absolute error perspective, the expected error in an aggregated estimate is smaller than the average of individual estimates unless all estimates are biased in the same direction (in which case the two are the same size).

Poll aggregators also must consider the scope of the surveys that will be averaged. Because survey results surrounding elections vary in response to campaign events (Erikson and Wlezien 2012b; Sides and Vavreck 2013), old polls may reflect information that is no longer current. Defining what time period is in scope represents a trade-off between statistical precision and survey relevance. Aggregations can also be generated at either the state or national level. The process of determining which surveys are worth averaging in the first place is a similar quandary. Some survey firms collect data from panels of respondents who sign up to take surveys in exchange for prizes (Baker et al. 2010); other firms rely principally or entirely on IVR surveys, which are typically prohibited from calling US mobile phones (Lavrakas et al. 2007); others are explicitly partisan; and many utilize proprietary strategies for weighting their data and defining which individuals are likely to vote (Voss, Gelman, and King 1995).

Different surveys may vary in their accuracy, representativeness, and even in the populations they describe. Any given aggregator must determine which surveys to include and whether they should be given equal weight in the model. Additionally, some survey outfits release multiple estimates from each survey. Choosing among these estimates can potentially alter an aggregator's conclusions. For example, surveys that did not include mobile phones reported results that demonstrated a pro-Republican bias in 2008 (Mokrzycki, Keeter, and Kennedy 2009).

Aggregated surveys are produced by a number of outlets and utilize a variety of methodological choices; the specific choices in any particular model are often opaque, however. CNN produces its "Poll of Polls" by averaging together three or more recent surveys on a particular topic that meet CNN's "methodological standards" (CNN n.d.). The website Electoral-vote.com averages vote preferences from the most recent presidential poll in each state with results from all other nonpartisan surveys conducted within the preceding week (Tanenbaum n.d.). RealClearPolitics.com appears to adopt a similar strategy for its polling averages, but no standard method for aggregating the polls is presented; notably, the time window for included polls and the set of surveys averaged seemed to vary across contests and states. Wang (2013) uses the median of all recent polls to produce state-level estimates for the Princeton Election Consortium.[5]

To date, there has been little clear evidence for which aggregation strategy is best. We lack both a systematic examination of aggregation strategies and clear benchmarks before election day for which individual survey results can be compared. Such an examination would behoove those hoping to compare different strategies for using these tools. At a minimum, however, aggregation is likely to be the most accurate to the extent that the process serves to mitigate likely biases. Hence, strategies combining estimates across multiple firms that

5. A simulation is made from these results to estimate electoral vote probabilities.

employ diverse methods and survey modes and that account for the precision of individual estimates are likely to provide the most accurate assessments. At present, the best metric we have for the precision of individual surveys is the historical accuracy of their final pre-election estimates.[6]

ADVANTAGES AND DISADVANTAGES OF AGGREGATION

Aggregation across surveys provides a number of distinct advantages over using a single survey estimate. Aggregating produces an estimate of the state of a contest that is less variable than individual survey results and less susceptible to idiosyncrasies that might skew any given estimate. Because aggregated models have greater precision than individual surveys, they can detect somewhat smaller changes in election dynamics (Erikson and Wlezien 1999).[7]

Yet, in many cases, naïvely aggregating across surveys can produce misleading results. Aggregators who weight all surveys equally ignore differences in survey sample size and quality (though it may be impossible to know which results are the highest quality or will yield the most accurate predictions of final election tallies a priori). Most aggregators do not account for the possibility of mode or house effects, whereby some survey techniques and firms can produce conclusions that are consistently different from the estimates of others (McDermott and Frankovic 2003; Panagopoulos 2009). One implication of this omission is that changes in the set of surveys considered might result in the incorrect conclusion that the dynamics of a contest had shifted. Ignoring potential sources of bias in polls when aggregating can also lead to artificially precise results, as the true error in a survey is generally larger than the sampling error (Crespi 1988).

Overall, aggregating polls has benefits over assessing each survey individually, but the aggregated estimate is not a perfect tool. Aggregations continue to describe the current state of an election rather than expectations for the final tallies and only address the many potential sources of error in surveys with the assumption that those errors will wash out in the aggregated estimate. Notably, this expectation depends on the existence multiple polls; hence, in many individual states, aggregation simply isn't possible. Further, any systematic difference between the set of surveys aggregated and the final election tallies remains.

6. Metrics such as sampling quality and sample size would ideally help, but the potential for systematic errors as well as the typically proprietary nature of likely voter models may limit inferences that can be drawn solely from survey design (cf. Crespi 1988; Jackman 2012). Unfortunately, there is evidence that the final pre-election estimates from survey houses are often manipulated, undermining the assumption that historical accuracy will predict future accuracy (Blumenthal 2014).

7. Since doubling the sample size reduces the margin of sampling error by almost 30 percent, the gains from aggregating can be substantial. Indeed, the standard error of an estimate is inversely related to the square of the sample size; doubling the size of the sample is thus equivalent to multiplying the standard error by $1/\sqrt{(2)}$, or approximately .707.

Predictive Modeling

Among the central concerns with election surveys is the focus they place on the momentary state of an election instead of the expected results on election day (Erikson, Panagopoulos, and Wlezien 2004; Moore 2008). Current dynamics are far more variable than election day projections (Gelman and King 1993). We know, for instance, that candidates tend to receive a bump in the polls at certain seminal moments: when they emerge from the primaries, during their party conventions, and—for some challengers—following the first debate (Erikson and Wlezien 2012b). Yet, many of these appear to be short-term dynamics (Sides and Vavreck 2013). Early poll results, in particular, tend to present a poor and overly variable portrait of an election, whereas election day tallies are better predicted from a set of principally economic measures, the so-called "fundamentals" (Gelman and King 1993; Wlezien and Erikson 2004; Campbell 2005). This is not to say that there is no useful information in election surveys; instead, the challenge is disentangling substantive indicators of electoral dynamics from the transitory tug-of-war of competing campaigns (Sides and Vavreck 2013).

Prediction models allow analysts to glean useful information from surveys and other data sources to enable election day forecasting. By accounting for how polls are expected to relate to voter behavior and discounting temporary shifts, poll-based forecasting models can mitigate the tendency for survey results to overreact to campaign events (Erikson and Wlezien 2012b).

To generate predictive models, researchers typically regress vote shares from previous elections onto a combination of polling averages and other predictors. Parameters from these models are then used to estimate vote shares in the current election. Covariates are typically assigned to control for theoretically relevant drivers of electoral behavior and for exogenous sources of error, such as survey house effects.[8] The result is a model that accounts for known differences between current observations and election expectations.

CHOICES IN PREDICTIVE MODELING

Developing prediction models requires a look to the past. Key variables that might help translate between polls and election results are found by considering the dynamics of previous contests. Researchers have explored performance in primary elections (Norpoth and Bednarczuk 2012), unemployment (Lewis-Beck and Tien 2008), and a tendency for incumbents to lose ground (Campbell and Wink 1990). Yet, not every potential election indicator can be examined. Each election result emerges from a unique constellation of circumstances that will never be replicated; no model can capture the impact of Howard Dean's "scream," Michael Dukakis's tank ride, Mitt Romney's "47%," or George

8. Controlling for house effects allows researchers to ignore variations over time that accrue from changes in the set of surveys that are assessed.

H. W. Bush's penchant for checking his watch during a town hall debate (cf. Shaw 1999). These idiosyncratic features of particular contests are useless as a predictor of dynamics in future elections and may not have much impact despite considerable media attention (cf. Gelman and King 1993). Instead, models must focus on more general patterns that emerge. A predictive model can project current survey results onto election day by focusing on common election dynamics or the tendency for survey information to converge with economically oriented projections of the election.

Determining what information should be included is the central challenge when using models to predict election results. The forecasting models that use surveys typically focus on a small number of political and economic variables that can be assessed well before election day (Campbell 2008, 2012). This parsimony ensures that prediction models provide strong tests of voting theory, but also serves to limit their explanatory power. Although a solid theoretical understanding should lead to accurate predictions, it is likely that many factors contribute in small but substantive ways to the eventual results. Some predictive models combine surveys with long-term aggregate trends in electoral preferences (e.g., Holbrook and DeSart 1999); others are designed to place polling results within the confines of general tendencies within an electoral cycle (e.g., Erikson and Wlezien 2012b). Because forecasting estimates consider the dynamics of past elections and incorporate elements of voting theory, they tend to outperform raw polling averages in predicting outcomes (cf. Erikson and Wlezien 2008a).

Most prediction models are the domain of political science. Poll-dependent predictive models are a subset of a larger group of forecasting models based on theories of vote choice. Among the models printed regularly in political science journals, candidate preferences are centrally used in Campbell's (1996) Trial Heat model and in Erikson and Wlezien's (2008b) model of Leading Economic Indicators and the Polls. These models were both within one percentage point of the national popular vote in 2012, despite forecasts made months in advance (Campbell 2012). Because prediction models are designed to test theories, the best models are those that highlight a single theoretical approach, test it parsimoniously, and provide accurate prospective estimates (see Lewis-Beck 2005).

Studies of prediction models have shown that survey estimates become increasingly accurate over the campaign cycle, diminishing the added value of regression-based corrections and additional indicators as the election draws near (Erikson and Wlezien 2008b). Predictive models are also produced at both the state (e.g., Campbell 1992; DeSart and Holbrook 2003) and national (e.g., Campbell and Wink 1990; Brown and Chappell 1999; Erikson and Wlezien 2008b) levels.[9]

9. The most prominent example of a political science forecasting model—the *Washington Post Wonkblog* model produced by Sides, Vavreck, and Hill—did not include polling data in 2012 (see Sides 2014).

ADVANTAGES AND DISADVANTAGES OF PREDICTIVE MODELS

Predictive models present two distinct advantages for estimation that are not present in simple aggregations. For one, the estimate they provide is designed to capture the election outcome. This may differ in systematic ways from the more fluid consensus of current polling. For another, predictive models can incorporate non-survey sources of information. In American elections, the final results frequently converge around a series of exogenous economic indicators. Scholars have argued that elections constitute a learning process whereby individuals gather sufficient information to vote in line with the economic portrait of the nation (Gelman and King 1993). Empirically, final tallies can often be estimated as a mixture of current sentiment and these economic fundamentals (Campbell 1996). Incorporating auxiliary information along with survey data thus tends to improve election prediction (Bélanger and Soroka 2012; Graefe et al. 2014a).

Specifying prediction models of elections can introduce complications, however. For large national elections such as the US presidency, inferences about relevant predictors can typically only be drawn from the handful of prior cases.[10] This severely constrains the potential complexity of national-level models. Because prediction models are built on only a few cases, the presence of a large number of covariates can lead to overfitting. Although models at the state level appear to have much larger databases, state-level estimates suffer from considerable autocorrelations, constraining their effective sample size (Jackman 2014). Also, fewer covariates are available at the state level, limiting the range of theories that these models can examine. The intersection of a theory-testing agenda with a limited learning database has led researchers working in the forecasting arena to emphasize model parsimony (Lewis-Beck 2005).[11]

To the extent that researchers are interested in predicting election outcomes, regression-based estimations tend to improve considerably on both the results of individual surveys and aggregations across surveys. The predictions they produce are theoretically grounded and empirically rigorous. Yet, the restricted capacity of these models understates our full ability to predict election results. Such estimates tend to include at most a few variables and are

10. There have only been 56 US presidential elections, and data for many of the variables of interest has been available for fewer than 20 of these. Hence, many models of electoral outcomes should improve notably as additional data becomes available for these events.

11. Notably, the predictive approach depends on a researcher-theorized model for estimation rather than on the design of the data-collection procedure. This offers advantages for theory testing, but also changes the nature of the inferences that are being made. Instead of relying on the construction of the sample as the rationale for an estimate (design-based inference), the believability of the estimate depends on the credibility of the model (model-based inference; Little 2004). Although predictive models are derived theoretically and their accuracy can be demonstrated empirically, results will depend on the set of predictors included, making their estimates subject, at least in part, to the researcher's assumptions.

therefore constrained in their ability to capture the full scope of influences in an election campaign.

Hybrid Modeling

Even before the first election poll, we already know lots of information about the upcoming election environment: American elections are overwhelmingly contested by two parties (cf. Lacy and Burden 1999); they tend to be concentrated in a handful of "battleground" states (Hill and McKee 2005); most people and states vote the same way year after year (Converse 1966; Jackman 2014); and even the largest landslides are unlikely to differentiate the parties by an enormous margin (Dahl 1990; Silver 2008). It seems silly to ignore all of this information when assessing the current state of an election or producing estimates of the election result.

As the campaign unfolds, we continue to learn about attitudes toward the candidates, state-level dynamics, and the economic backdrop for the election. Ideally, we would like to integrate newly emerging information and assessments into our understanding of the current state of play. Poll aggregation strategies are unable to assess these dynamics because they concentrate only on a small time window and do not have a mechanism for incorporating additional information. Predictive models are constrained by their focus on a small set of predictors often derived from only a handful of elections. In contrast, hybrid modeling allows researchers to simultaneously incorporate the results from the preexisting electoral environment, emerging election polls, economic data, and other sources of information in a manner that is true to what is—and is not—known about an election.

At its heart, hybrid modeling is any strategy for combining diverse approaches into a single set of estimates. The simplest hybrid modeling technique involves pooling a series of estimates to generate a new forecast that incorporates elements from each model. This approach has been used in the aggregated forecasting estimates generated by Campbell (2008, 2012) and by Pollyvote (Graefe et al. 2014b). Theoretically, it can be improved upon by weighting each predictor or forecast by the precision with which it was generated (Montgomery, Hollenbach, and Ward 2012). In this manner, hybrid approaches triangulate information across multiple estimation strategies.

To incorporate a wide variety of predictors, hybrid approaches typically estimate models using diverse parameters and then provide a formula for combining those models. There are two basic approaches to producing hybrid estimates: a standard approach, wherein researchers define how the models should be combined, and a Bayesian approach, whereby the predictive precision of each estimate is itself parameterized and estimated as the model is run. The Pollyvote model is an example of the standard approach, whereby concurrent estimates from polls, experts, forecasting models, and prediction markets are averaged to generate election estimates (Graefe et al. 2014b).

Bayesian hybrid models offer a few advantages over the standard modeling approach. In particular, these models can include unique parameters for how national polling might be expected to relate to the polls in any given state and how state polls might relate to one another. A separate set of parameters might be estimated for how national and state polls would each be expected to relate to economic "fundamentals" (Lock and Gelman 2010; Linzer 2013a). Each new piece of data gathered during the course of the campaign can then be used to better estimate all related parameters. For example, a new Pennsylvania poll will increase information about how Pennsylvania relates to national survey results as well as to other states (Linzer 2013a). When national polls shift, the parameters of this sort of Bayesian hybrid model can then be used to project a corresponding change in Pennsylvania, even if no new data are available from the state. Further, parameters can be included that account for potential systematic differences between survey firms and survey modes, allowing researchers to create comparable trends over time (e.g., Linzer 2013a).

CHOICES IN HYBRID MODELING

Although the principles of generating a hybrid model are straightforward, estimating one is often considerably more complex. To produce the most accurate possible model, researchers need to properly parameterize not only the relations between the predictors that are included, but also the uncertainty of those relations as well as the precision of each of the measurements (Jackman 2004). Estimation thus requires two notable steps. First, researchers must start with some set of expectations for each parameter to be estimated; in the Bayesian context, this is known as a prior. To initialize a model, prior estimates are produced to account for what is known about each parameter to be estimated; frequently, these initial expectations are considered highly uncertain so that new data will drive the model (Jackman 2004). For standard hybrid models, all estimators are often treated as equally predictive (see Graefe et al. [2014b] for a discussion). Second, to understand what is known about any given parameter, the priors can be updated using data gathered from a survey or any other source, known in the Bayesian context as a likelihood, to produce a posterior estimate (Gill 2002). Posterior estimates are generated as a precision-weighted average of prior and likelihood. Standard hybrid models can apply parameters estimated from earlier elections to update estimates of the precision of each approach, whereas Bayesian estimates are updated using Markov chain Monte Carlo methods (Jackman 2004). The use of Monte Carlo simulations ensures that uncertainty in the *posterior* parameters is properly reflected in the final estimates.

Three choices guide the hybrid modeler; central decisions concern what sources of information should be included, what relations should be parameterized, and how the values of the priors should be initialized (Christensen and Florence 2008).[12] Although the same types of additional variables can

12. Bayesian models are also influenced by the functional form of priors (Lavine 1991).

be evaluated, considerations about what information should be included differ between a Bayesian hybrid model and a more traditional hybrid model. Because Bayesian models use Monte Carlo simulation—and thereby estimate out-of-sample error—there is little concern that the inclusion of additional predictors will derail a well-specified model (Gill 2002).[13] For standard models, poor predictors will tend to introduce additional random error.

The chief constraint on the number of predictors that can be included in a Bayesian model is the time it takes to run the model, which is frequently substantial and grows as the number of parameters increases. Weakly related estimators thus increase model run time for little substantive gain. Parameterized relations are referred to in a Bayesian context as the structural model. As with sources of information, adding additional parameters to be estimated increases the time required to estimate the model, but should not lead to inferential errors. Finally, there are two common strategies for setting initial specifications in a Bayesian model. Noninformative priors assume that almost nothing is initially known about the likely state of an election; in contrast, informative priors incorporate information about the likely status of an election, but may bias results toward historical outcomes (Rigdon et al. 2009).

Nate Silver's FiveThirtyEight is the most prominent example of hybrid modeling as a way to forecast election results. Although a complete methodological description is not available, Silver reports a sophisticated approach that incorporates state- and national-level surveys, a decay function that downweights the relevance of older polls, sample-size-based weighting, regression to the mean difference between the candidates, estimated differences between registered and likely voter surveys, and weights to account for house effects.[14] The FiveThirtyEight model also reports the use of "state fundamentals," including partisanship, candidate fund-raising, political ideology, and candidate quality measures (Silver n.d.). These two estimation strategies appear to be incorporated into measures of both the current state of the election and a state-by-state prediction for election day. It is unclear whether the combination of factors used for the FiveThirtyEight model is itself parameterized (which would make it a Bayesian model) or whether these are instead fixed (rendering it a more traditional hybrid model).

Simon Jackman's model, used by Pollster in 2012, incorporates state and national polling into a more explicitly Bayesian framework; the model clusters politically similar states and uses new polling information to update past polling estimates (rather than assigning a weight to specific older polls; see Jackman and Blumenthal [2013]). A more complex model is utilized by Linzer

13. So long as parameters have priors that properly index what researchers know (both expectations and uncertainty) about a presumed set of relations (cf. van Dongen 2006).

14. Because full information on the FiveThirtyEight model is not available online, it is not clear whether the error analysis used in the model is truly Bayesian. Given the way that Silver references his model elsewhere and the description in Silver (n.d.) of his 2010 senatorial model, the use of a hybrid approach is clear.

(2013a, 2013b) for Votamatic, which also incorporates historical results and economic "fundamentals." By the end of the election, model parameters and estimates are continuously updated from a running series of newly published polls (Jackman 2005; Linzer 2013a).

ADVANTAGES AND DISADVANTAGES OF HYBRID MODELING

Hybrid models have advantages over other strategies for pooling surveys in the information they can incorporate. Hybrid models can simultaneously leverage polls, fundamentals, historical performance, betting markets, and expert ratings to generate estimates. Some can even borrow statistical power from polls in one state to better estimate the likely outcome in another state (Lock and Gelman 2010; Jackman and Blumenthal 2013). Further, these approaches allow us to consider relations between all of the relevant factors in our model; with Bayesian estimation, it is possible to continuously triangulate information across numerous predictors. Together, these properties let us generate simultaneous state- and federal-level estimates and allow for predictions of out-of-sample cases—such as un-polled states or election day projections—that are more accurate than would be possible with other methods (Linzer 2013a).

A central challenge with hybrid models is their complexity. These models generate estimates by working between many data sources. As noted above, priors must be assigned to each of the parameters to be estimated. If inappropriate priors are assigned, models can reach erroneous conclusions. For example, putting too strong a weight on past performance in Virginia in 2008 would have led a model astray. Unfortunately, readers often encounter these models without full knowledge of the priors that were assigned, the functional form of those priors, or the structural model within which those priors were nested. Without this information, consumers can hardly evaluate the appropriateness of the model.[15] As an example, Jackman's model works from the assumption that surveys are, on average, unbiased estimators of the state of a contest (Jackman and Blumenthal 2013). Although this constraint makes estimation easier, it may or may not be accurate (it also means that the Pollster model should be interpreted as a reflection of the state of affairs instead of as an election day prediction; see Pickup and Johnston [2008]). Bayesian models, in particular, also have large computational requirements (both in programming skill and run time), which limit their accessibility to many researchers. Finally, these models are data driven rather than solely theory driven. Although structural models and data collection may be specified with theory in mind, the goal

15. Wildly inappropriate assumptions for hybrid models should become apparent when consumers are confronted with estimates that do not seem credible. Notably, our understanding of what went wrong in failed forecasts tends to be considerably stronger for predictive models than for hybrid models, because we can understand the sources of error. For standard hybrid models, it is often easy to see which component led to an error, but the reasons are often still opaque.

is prediction rather than a substantive understanding of the electoral dynamics at play.

Altogether, hybrid models have produced the most efficient and accurate estimates of recent election outcomes (Muehlhauser and Branwen 2012; Graefe et al. 2014b; Branwen n.d.). The process is in many ways ideal for a complex problem like election forecasting, where many substantive variables are relevant for understanding elections and only a handful of prior elections can be used for guidance. The procedure is sensitive to model assumptions and specifications, however, and modeling is sufficiently complex that the full set of predictors and procedures is often opaque to many of those who wish to understand contemporary elections. Hence, despite recent predictive success and efficiency advantages in incorporating more sources of data into estimation, the applications of hybrid models are limited.

Comparing Pooling Strategies

To illustrate the relative usefulness of the various methods of combining surveys, nine data streams were compared in their ability to prospectively predict election day results in 2012 and to track the election.[16] Differences between election day tallies and estimates from each of nine pooling approaches were generated for a 75-day window from August 23 to November 6, 2012, a period that included both party conventions and all three presidential debates. Where possible, estimates were generated at both the state and national levels. The nine estimation methods compared are shown in table 1; additional description on how all estimates were calculated is provided in the online appendix.

Because different pooling approaches are designed to accomplish different goals, three metrics were generated for each of the methods; these were then compared to similar metrics for the other types of approaches. Methods are compared in the predictive accuracy of their stream of estimates (where each day's estimate is treated as a forecast), the accuracy of their final estimates, and the responsiveness of methods to variations in the campaign. In general, methods designed for forecasting the election should be evaluated by their ability to prospectively estimate election day results, whereas methods designed for tracking the campaign should be evaluated by their responsiveness to campaign events. Both goals should presumably lead to accurate estimates just before election day.

Because different methods provided estimates for differing subsets of states and dates (which would be expected to have differing levels of precision), we needed to place these methods on an equivalent playing field. Otherwise, we might penalize some models for errors in states like West Virginia, where there were few polls, while lauding models that did not even attempt to estimate the

16. Election day tallies came from each state's Certificate of Ascertainment. These were summed to produce vote totals for the national election.

Table 1. Comparison of Models Predicting 2012 Results

Model	Type	Goal	Variables included	Generation strategy
Avg. recent survey	N/A	Tracking	- Survey results	Mean error of surveys conducted in 7-day rolling window.
Aggregated recent surveys	Survey aggregation	Tracking	- Survey results	Mean of surveys conducted in 7-day rolling window
RealClearPolitics[a,b]	Survey aggregation	Tracking	- Survey results	Full methodology not available; national data collected daily by Graefe (2013)
Erikson & Wlezien (2012a)[a]	Prediction model	Forecast	- Survey results (national surveys) - Cumulative leading economic indicator index (LEI)	Projected national vote share = 19.68 + 4.08 (LEI = .34) + .59 (national surveys)
DeSart & Holbrook (2009)	Prediction model	Forecast	- Survey results (state surveys) - Survey results (national surveys) - Two-party performance in prior 4 elections (prior vote)	Projected state vote share = –13.906 + .75 (state surveys) + .364 (prior vote) + .163 (national surveys)
Jackman (Pollster)[c]	Hybrid Bayesian model	Tracking	- Survey results (state and national)	1) Define a parameter for how each pair of states relates. 2) Estimate parameters for all states with polls. 3) Estimate house effects for each polling firm and precision for each poll. 4) Update state estimates using these parameters when each new poll (in any state or nationally) comes out.

Continued

Table 1. *Continued*

Model	Type	Goal	Variables included	Generation strategy
Graefe et al. (PollyVote)[a,b]	Hybrid traditional model	Forecast	- Survey aggregations - Prediction markets - Expert forecasts - Predictive models - Index models	1) Average all estimates from each model type. 2) Average estimates across all model types.
Silver (FiveThirtyEight)[a,b]	Hybrid model (type unknown)	Forecast	- Survey results (state and national) - Partisanship/ideology - Candidate fund-raising - Candidate quality	Full methodology not available; national data collected daily by Graefe (2013).
Linzer (Votamatic)[d]	Hybrid Bayesian model	Forecast	- Survey results (state and national) - Two-party performance in prior election - Historical trends	1) Start with historical performance of each state. 2) Treat each state poll as the result of a mixture of state and national factors. 3) Consider election day to be a random walk from the most recent estimate of the model influenced by known trends from prediction models and history.

NOTE.—Unless otherwise noted, estimates for data streams were generated by the author using data from pollster's free public API (http://elections.huffington-post.com/pollster/api). Although Campbell and Wink's (1990) trial-heat model also uses survey data, there is no mechanism for updating the results with newer survey data as the election approaches; hence, it was excluded from the current analysis.

[a]Data available only for national comparisons.

[b]Data provided by Graefe (2013).

[c]Data provided by Mark Blumenthal.

[d]State-level estimates were provided by Drew Linzer; national estimates for this model were produced as a weighted average of state-level estimates. In order to maintain a prospective estimate, these averages were generated using the proportion of the national vote generated by each state in the 2008 election.

state. To address this concern, all comparisons between methods were assessed at the dyad level. Dyads were defined as the set of unique date-state combinations that could be matched across any given pair of methods for any given analysis.

For assessing the predictive accuracy of estimates from separate models, dyadic cases were used to compare a pair of methods that provided estimates in a given state (or nationally) on a given day. Table 2 presents the proportion of all comparable state-day combinations for which the method in the rows produced an estimate closer to official final election tallies than did all methods of the type listed in the columns.[17] For example, compared to the average of recent surveys, the RealClearPolitics estimate was closer to final election tallies 93.7 percent of the time (row 3, column 1). The better prediction from RealClearPolitics was significantly different from chance, as assessed by a binomial test ($p < .001$).

All approaches to pooling appear to provide a significant benefit over the use of a single recent survey for predicting future election outcomes. Estimates from the various methods provided more accurate forecasts than un-pooled survey results from 73.5 to 100.0 percent of the time (table 2, column 1). Election forecasts from predictive and hybrid models were also consistently more accurate than those using aggregation; these more advanced methods outperformed aggregation for between 60.8 and 80.0 percent of cases, depending on the method used (column 2). Comparisons between models depended largely on the goals of the pooling strategy; the Jackman model, which was designed for tracking rather than forecasting the election, had the smallest relative advantage over the aggregation methods and did not seem to perform as well as the predictive models or other hybrid models. Among the approaches designed for forecasting the election, the Linzer and Silver hybrid models yielded the best estimates, outperforming predictive models (column 3) as well as other hybrid approaches (column 4).

Limiting the set of comparisons to estimates produced on November 5, 2012, yielded similar results (table 3).[18] The dyads described in table 3 thus include only the final pre-election estimate for each method from each state. Again, all of the methods outperformed the use of a typical recent survey (column 1). Predictive models and hybrid models provided better election day forecasts than aggregation methods (column 2). And hybrid models tended to perform better than predictive models (column 3). The notable exception was for the hybrid model designed to track the election rather than forecast final election tallies (row 6); the inferior final performance of Jackman's model compared to the forecasting strategies is likely due to the decision to center estimates using the average of all available polls, which was slightly biased in the 2012 cycle (Jackman and Blumenthal 2013).

17. Cases where equivalent estimates were produced using both methods in a dyad were dropped.
18. Surveys used for these estimates had field dates ending between October 30 and November 5, using the same 7-day window as the method shown in table 2.

Table 2. Dyadic Comparisons of Model Accuracy for Estimates of Election Day Tallies Made over the 75 Days before Election Day

	Proportion of dyads for which method in row is more accurate than estimates from (other)									
	Avg. recent survey		*Aggregation methods*		*Predictive models*		*Hybrid models*		Total	
	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*
Avg. recent survey	–		0.3***	629	18.9***	859	22.3***	1702	17.1***	3190
Aggregation methods										
Aggregated recent surveys	100.0***	554	33.3**	75	29.5***	859	33.6***	1702	44.0***	3190
RealClearPolitics	97.3***	75	66.7**	75	28.8***	118	32.8***	238	46.4	506
Predictive models										
Erikson & Wlezien	100.0***	75	70.7***	150	30.2**	43	49.2	238	61.5***	506
DeSart & Holbrook	79.3***	784	70.6***	827	69.8**	43	50.7	1900	61.9***	3554
Hybrid models										
Jackman (Pollster)	73.5***	956	60.8***	1008	42.1***	1207	37.8***	1207	52.1**	4378
Graefe et al. (Pollyvote)	100.0***	75	70.0***	150	28.0***	118	36.8***	163	54.0	506
Linzer (Votamatic)	78.7***	596	71.5***	632	62.2***	695	61.3***	1175	66.9***	3098
Silver (FiveThirtyEight)	100.0***	75	80.0***	150	72.0***	118	72.4***	163	78.7***	506

NOTE.—*N*s are numbers of dyads (comparing estimates from two models in a single state on a single day) available for comparisons between model in rows and all models of the types listed in columns, with equal estimates dropped. Statistical significance was calculated using binomial tests.

$^{**}p < .01$; $^{***}p < .001$, two-tailed difference from 50 percent.

Table 3. Dyadic Comparisons of Model Accuracy for Final Estimates of Election Day Tallies

	Proportion of dyads for which method in row is more accurate than estimates from (other)									
	Avg. recent survey		*Aggregation methods*		*Predictive models*		*Hybrid models*		Total	
	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*
Avg. recent survey	–		0.0***	17	25.9**	27	22.2***	54	19.4***	98
Aggregation methods										
Aggregated recent surveys	100.0***	16	100.0[a]	1	33.3	27	29.6**	54	42.9	98
RealClearPolitics	100.0[a]	1	0.0[a]	1	0.0[a]	2	0.0[a]	4	12.5**	8
Predictive models										
Erikson & Wlezien	100.0[a]	1	100.0[a]	2	100.0[a]	1	50.0	4	75.0	8
Desart & Holbrook	73.1**	26	66.7	27	0.0[a]	1	49.3	75	57.4	129
Hybrid models										
Jackman (Pollster)	73.1**	26	63.0	27	36.8	38	29.5**	44	46.7	135
Graefe et al. (Pollyvote)	100.0[a]	1	100.0[a]	2	50.0	2	0.0[a]	3	50.0	8
Linzer (Votamatic)	80.8***	26	77.8**	27	62.2	37	72.7**	44	72.4***	134
Silver (FiveThirtyEight)	100.0[a]	1	100.0[a]	2	100.0[a]	2	66.7	3	87.5**	8

NOTE.—*N*s are numbers of dyads (comparing final estimates from two models in a single state) available for comparisons between model in rows and all models of the types listed in columns, with equal estimates dropped. Statistical significance was calculated using binomial tests.

$^{**}p < .01$; $^{***}p < .001$, two-tailed difference from 50 percent.

[a]Statistical significance is not shown for comparisons with fewer than 8 cases if no variation is present, as these would not be significant with a single countervailing case.

Finally, I assessed the responsiveness of the pooling strategies to changing election dynamics. Because identifying central campaign events in response to polling fluctuations would be akin to selecting on the dependent variable, responsiveness was parameterized as the overall variance in the estimates provided for any given state-method combination.[19] For each dyadic comparison, the variance for Obama's estimated proportion of two-party votes in each state was calculated for a pair of methods using the set of shared dates for which both methods provided estimates. The proportion of cases where each method had a larger variance, and thus was presumed to vary more in response to electoral events, is reported in table 4.

The variance of estimates from the different methods presented an inverse pattern to their forecasting accuracy. Separate surveys varied more than did survey aggregations, predictive models, or hybrid models (table 4, row 1). The aggregation strategies were typically more variable than both the predictive and hybrid models (column 2). Among modeled approaches, the Jackman estimates were the most variable, in line with claims that the model was tracking the election rather than forecasting its results (row 6). Finally, the two hybrid models that provided the most accurate overall estimates also presented the least variable portraits of the last few months of the election.

Results presented here illustrate that the prediction and hybrid models tended to provide the most accurate forecasting estimates preceding the 2012 election. The best of these predictors were hybrid models designed to predict election day behaviors. Of course, little should be made from the advantageous performance of only two models over a single election cycle, but the findings are in line with the theoretical advantages for this type of approach. Similar comparisons by Graefe et al. (2014a, 2014b) indicate benefits from combining forecasts into a hybrid approach for even longer-term forecasting. The results of another comparison of final pre-election estimates in each state and nationally also revealed that hybrid modeling provided the most accurate estimates (Branwen n.d.).

Pooling the Polls

Election surveys have come a long way in the past century; our methods for conducting them have improved exponentially, and our tools for combining and understanding them have encountered similar innovations. Yet, deciding which strategy to use when pooling surveys is not simply a matter of running the most complex and accurate model. These choices should instead be guided by the questions we wish to ask and the phenomena we hope to understand; a desire for simplicity or theory testing may sometimes trump the most accurate predictions.

19. Treating party conventions and presidential debates as predictors in a regression discontinuity model produced similar results, but was not shown.

Table 4. Dyadic Comparisons of Variance for Daily Estimates by State

	Proportion of dyads for which method in row is more variable than estimates from (other)									
	Avg. recent survey		*Aggregation methods*		*Predictive models*		*Hybrid models*		Total	
	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*
Avg. recent survey	–		95.0***	20	100.0***	27	96.4***	55	97.1***	102
Aggregation methods										
Aggregated recent surveys	5.3***	19	100.0[a]	1	89.2***	37	80.5***	77	72.4***	134
RealClearPolitics	0.0[a]	1	0.0[a]	1	100.0[a]	2	75.0	4	62.5	8
Predictive models										
Erikson & Wlezien	0.0[a]	1	0.0[a]	2	0.0[a]	1	50.0	4	25.0	8
DeSart & Holbrook	0.0[a]	26	10.8***	37	100.0[a]	1	35.6**	73	22.6***	137
Hybrid models										
Jackman (Pollster)	3.1***	32	21.6***	37	83.8***	37	95.5***	44	54.7	150
Graefe et al. (Pollyvote)	0.0[a]	1	0.0[a]	2	100.0[a]	2	66.7	3	50.0	8
Linzer (Votamatic)	4.8***	21	20.0***	40	44.4	36	4.5***	44	19.1***	141
Silver (FiveThirtyEight)	0.0[a]	1	0.0[a]	2	0.0[a]	2	33.3	3	12.5**	8

NOTE.—*N*s are numbers of dyads (comparing variance over shared dates from two models in a single state) available for comparisons between model in rows and all models of the types listed in columns, with equal variances dropped. Statistical significance was calculated using binomial tests.

$^{**}p < .01$, $^{***}p < .001$, two-tailed difference from 50 percent.

[a]Statistical significance is not shown for comparisons with fewer than 8 cases if no variation is present, as these would not be significant with a single countervailing case.

Even with today's advanced statistical tool kit, the simple aggregation of survey results has an important role to play in our understanding of elections. With the preponderance of survey data available, aggregation estimates are a quick and efficient means of gleaning the current state of an election contest. They provide a window that highlights the day's new developments and tracks the campaign horse race in an accessible and digestible manner.[20] They also allow us to assess daily variations that are smaller than what can be captured by a single survey.

Predictive election models are also important because they provide meaningful insight into what we understand about election campaigns and their eventual results. Constrained by their parsimony, these models nonetheless tell us enormous amounts about the likely results on election day and allow us to test theoretical claims about the core variables that drive electoral decisions and election results. Their predictive success provides a key barometer for what we do and do not comprehend about how elections work.

Finally, hybrid models have the potential to provide the most accurate estimates of eventual election tallies. They allow us to combine our insights to produce meaningful, accurate, and useful prospective predictions. When the methods are made available, they also allow us to gain a substantive understanding of the interplay between the various electoral forces at work by revealing which of our predictors was most accurate. Unfortunately, this information is too often unavailable. If our purpose is projecting electoral outcomes, these are, at least in theory, the strongest tools we have in our current arsenal.

Researchers choosing a strategy for averaging the polls need to consider what they hope to understand. The right model depends on whether the message is about dynamic change in the electorate, key theoretical indicators of election behavior, or producing the most accurate possible prediction.

Supplementary Data

Supplementary data are freely available online at http://poq.oxfordjournals.org/. Replication code is available at http://joshpasek.com/replication-files/replication-code-for-predicting-elections/.

References

Abrams, Mark. 1970. "The Opinion Polls and the 1970 British General Election." *Public Opinion Quarterly* 34:317–24.

Armstrong, J. Scott. 2001. "Combining Forecasts." In *Principles of Forecasting: A Handbook for Researchers and Practitioners*, edited by J. Scott Armstrong, 417–40. Hingham, MA: Kluwer Academic Publishers.

20. Whether campaign coverage *should* be focused on the horse race or not is an entirely different matter. A summary of the academic work on that question can be found in Iyengar, Norpoth, and Hahn (2004).

Baker, Reg, Stephen J. Blumberg, J. Michael Brick, Mick P. Couper, Melanie Courtright, J. Michael Dennis, Don Dillman, Martin R. Frankel, Philip Garland, Robert M. Groves, Courtney Kennedy, Jon Krosnick, and Paul J. Lavrakas. 2010. "AAPOR Report on Online Panels." *Public Opinion Quarterly* 74:711–81.

Bartels, Larry M. 1996. "Pooling Disparate Observations." *American Journal of Political Science* 40:905–42.

Bélanger, Éric, and Stuart N. Soroka. 2012. "Campaigns and the Prediction of Election Outcomes: Can Historical and Campaign-Period Prediction Models Be Combined?" *Electoral Studies* 31:702–14.

Berg, Joyce E., and Thomas A. Rietz. 2014. "Market Design, Manipulation, and Accuracy in Political Prediction Markets: Lessons From the Iowa Electronic Markets." *PS: Political Science and Politics* 47:293–96.

Berinsky, Adam J. 2006. "American Public Opinion in the 1930s and 1940s." *Public Opinion Quarterly* 70(4):499–529.

Blumenthal, Mark M. 2014. "Polls, Forecasts, and Aggregators." *PS: Political Science and Politics* 47:297–300.

Bohm, Peter, and Joakim Sonnegard. 1999. "Political Stock Markets and Unreliable Polls." *Scandinavian Journal of Economics* 101:205–22.

Branwen, Gwern. "2012 Election Predictions." gwern.net. Available at http://www.gwern.net/2012%20election%20predictions.

Brown, Lloyd B., and Henry W. Chappell Jr. 1999. "Forecasting Presidential Elections Using History and Polls." *International Journal of Forecasting* 15:127–35.

Campbell, James E. 1992. "Forecasting the Presidential Vote in the States." *American Journal of Political Science* 36:386–407.

———. 1996. "Polls and Votes: The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns." *American Politics Research* 24:408–33.

———. 2004. "Introduction—The 2004 Presidential Election Forecasts." *PS: Political Science and Politics* 37:733–35.

———. 2005. "The Fundamentals in US Presidential Elections: Public Opinion, the Economy and Incumbency in the 2004 Presidential Election." *Journal of Elections, Public Opinion, and Parties* 15:73–83.

———. 2008. "Editor's Introduction: Forecasting the 2008 National Elections." *PS: Political Science and Politics* 41:1–4.

———. 2012. "Forecasting the 2012 American National Elections." *PS: Political Science and Politics* 45:610–13.

Campbell, James E., and Kenneth A. Wink. 1990. "Trial-Heat Forecasts of the Presidential Vote." *American Politics Research* 18:251–69.

Christensen, William F., and Lindsay W Florence. 2008. "Predicting Presidential and Other Multistage Election Outcomes Using State-Level Pre-Election Polls." *American Statistician* 62:1–10.

CNN. "CNN Poll of Polls: Key Battleground States." CNN. Available at http://www.cnn.com/interactive/2012/10/politics/poll.of.polls/.

Cochran, William Gemmell. 1953. *Sampling Techniques*. New York: Wiley.

Converse, Jean. 1987. *Survey Research in the United States: Roots and Emergence 1890–1960*. Berkeley: University of California Press.

Converse, Philip E. 1966. "The Concept of a Normal Vote." In *Elections and the Political Order*, edited by Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes, 9–39. New York: Wiley.

Crespi, Irving. 1988. *Pre-Election Polling*. New York: Russell Sage Foundation.

Crossley, Archibald M. 1937. "Straw Polls in 1936." *Public Opinion Quarterly* 1:24–35.

Dahl, Robert Alan. 1990. "Myth of the Presidential Mandate." *Political Science Quarterly* 105:355–72.

DeSart, Jay A., and Thomas M. Holbrook. 2003. "Statewide Trial-Heat Polls and the 2000 Presidential Election: A Forecast Model." *New Directions for Program Evaluation* 84:561–73.

———. 2009. "A Tale of Two Models: The DeSart and Holbrook State-Level Forecast Model in 2008." Paper presented at the Annual Meeting of the Western Political Science Foundation, Vancouver, BC.

Erikson, Robert S., Costas Panagopoulos, and Christopher Wlezien. 2004. "Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics." *Public Opinion Quarterly* 68:588–601.

Erikson, Robert S., and Christopher Wlezien. 1999. "Presidential Polls as a Time Series: The Case of 1996." *Public Opinion Quarterly* 63:163–77.

——— 2008a. "Are Political Markets Really Superior to Polls as Election Predictors?" *Public Opinion Quarterly* 72:190–215.

———. 2008b. "Leading Economic Indicators, the Polls, and the Presidential Vote." *PS: Political Science and Politics* 41:703–07.

———. 2012a. "The Objective and Subjective Economy and the Presidential Vote." *PS: Political Science and Politics* 45:620–24.

———. 2012b. *The Timeline of Presidential Elections*. Chicago: University of Chicago Press.

Gallup, George. 1951. "The Gallup Poll and the 1950 Election." *Public Opinion Quarterly* 15:16–22.

Gelman, Andrew, and Gary King. 1993. "Why Are American Presidential Election Campaign Polls So Variable When Votes Are So Predictable?" *British Journal of Political Science* 23:409–51.

Gill, Jeff. 2002. *Bayesian Methods*. London: Chapman & Hall/CRC.

Graefe, Andreas. 2013. "Replication Data for Accuracy of Combined Forecasts for the 2012 Presidential Election: The PollyVote." Harvard Dataverse Network. Available at http://thedata.harvard.edu/dvn/faces/study/StudyPage.xhtml?globalId=doi:10.7910/DVN/POLLYVOTE2012&studyListingIndex=0_746e1b05d2d7d00f4c8959c473d6.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones Jr., and Alfred G. Cuzan. 2014a. "Accuracy of Combined Forecasts for the 2012 Presidential Elections: The Pollyvote." *PS: Political Science and Politics* 47:427–31.

———. 2014b. "Combining Forecasts: An Application to Elections." *International Journal of Forecasting* 30:43–54.

Green, Donald P., Alan S. Gerber, and Suzanna L. De Boef. 1999. "Tracking Opinion over Time: A Method for Reducing Sampling Error." *Public Opinion Quarterly* 63:178–92.

Groves, Robert M., and Lars Lyberg. 2011. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74:849–79.

Hansen, Jan, Carsten Schmidt, and Martin Strobel. 2004. "Manipulation in Political Stock Markets—Preconditions and Evidence." *Applied Economics Letters* 11:459–63.

Herbst, Susan. 1993. *Numbered Voices: How Opinion Polling Has Shaped American Politics*. Chicago: University of Chicago Press.

Hill, David B., and Seth C. McKee. 2005. "The Electoral College, Mobilization, and Turnout in the 2000 Presidential Election." *American Politics Research* 33:700–725.

Hillygus, D. Sunshine. 2011. "The Evolution of Election Polling in the United States." *Public Opinion Quarterly* 75(5):962–81.

Holbrook, Thomas M., and Jay A. DeSart. 1999. "Using State Polls to Forecast Presidential Election Outcomes in the American States." *International Journal of Forecasting* 15:137–42.

Iyengar, Shanto, Helmut Norpoth, and Kyu S. Hahn. 2004. "Consumer Demand for Election News: The Horserace Sells." *Journal of Politics* 66:157–75.

Jackman, Simon. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7:483–505.

———. 2005. "Pooling the Polls over an Election Campaign." *Australian Journal of Political Science* 40:499–517.

———. 2012. "House Effects, Back by Popular Demand." Huffpost Pollster. Available at http://www.huffingtonpost.com/simon-jackman/house-effects-by-back-by-_b_2007907.html.

———. 2014. "The Predictive Power of Uniform Swing." *PS: Political Science and Politics* 47:317–21.

Jackman, Simon, and Mark M. Blumenthal. 2013. "Using Model-Based Poll Averaging to Evaluate the 2012 Polls and Posters." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Boston, MA, USA.

Katz, Daniel. 1941. "The Public Opinion Polls and the 1940 Election." *Public Opinion Quarterly* 5:52–78.

Kou, S. G., and Michael E. Sobel. 2004. "Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls." *Political Analysis* 12:277–95.

Lacy, Dean, and Barry Burden. 1999. "The Vote-Stealing and Turnout Effects of Ross Perot in the 1992 US Presidential Election." *American Journal of Political Science* 43:233–55.

Lavine, Michael. 1991. "Sensitivity in Bayesian Statistics: The Prior and the Likelihood." *Journal of the American Statistical Association* 86(414):396–99.

Lavrakas, Paul J., Charles D. Shuttles, Charlotte Steeh, and Howard Fienberg. 2007. "The State of Surveying Cell Phone Numbers in the United States: 2007 and Beyond." *Public Opinion Quarterly* 71:840–54.

Lewis-Beck, Michael S. 2005. "Election Forecasting: Principles and Practice." *British Journal of Politics & International Relations* 7:145–64.

Lewis-Beck, Michael S., and Charles Tien. 2008. "The Job of President and the Jobs Model Forecast: Obama for '08?" *PS: Political Science and Politics* 41:1–4.

Linzer, Drew A. 2013a. "Dynamic Bayesian Forecasting of Presidential Elections in the States." *Journal of the American Statistical Association* 108(501):124–34.

———. 2013b. "How It Works | VOTAMATIC." votamatic.org. Available at http://votamatic.org/how-it-works/.

Little, Roderick J. A. 2004. "To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling." *Journal of the American Statistical Association* 99(466):546–56.

Lock, Keri, and Andrew Gelman. 2010. "Bayesian Combination of State Polls and Election Forecasts." *Political Analysis* 18:337–48.

McDermott, Monika L., and Kathleen A. Frankovic. 2003. "Review: Horserace Polling and Survey Method Effects: An Analysis of the 2000 Campaign." *Public Opinion Quarterly* 67:244–64.

Mitofsky, Warren J. 1998. "Review: Was 1996 a Worse Year for Polls Than 1948?" *Public Opinion Quarterly* 62:230–49.

Mokrzycki, Michael, Scott Keeter, and Courtney Kennedy. 2009. "Cell-Phone-Only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias." *Public Opinion Quarterly* 73:845–65.

Montgomery, Jacob M., Florian M. Hollenbach, and Michael D. Ward. 2012. "Ensemble Predictions of the 2012 US Presidential Election." *Political Science & Politics* 45:651–54.

Moore, David W. 2008. *The Opinion Makers: An Insider Exposes the Truth Behind the Polls*. Boston, MA: Beacon Press.

Mosteller, Frederick. 1948. "On Pooling Data." *Journal of the American Statistical Association* 43(242):231–42.

Muehlhauser, Luke, and Gwern Branwen. 2012. "Was Nate Silver the Most Accurate 2012 Election Pundit?" Center for Applied Rationality. Available at http://rationality.org/2012/11/09/was-nate-silver-the-most-accurate-2012-election-pundit/.

Norpoth, Helmut, and Michael Bednarczuk. 2012. "History and Primary: The Obama Reelection." *PS: Political Science and Politics* 45:614–17.

Panagopoulos, Costas. 2009. "Polls and Elections: Preelection Poll Accuracy in the 2008 General Elections." *Presidential Studies Quarterly* 39:896–907.

Pasek, Josh, and Jon A. Krosnick. 2010. "Optimizing Survey Questionnaire Design in Political Science: Insights From Psychology." In *The Oxford Handbook of American Elections and*

Political Behavior, edited by Jan E. Leighley and George C. Edwards III, 27–50. Oxford, UK: Oxford University Press.
Pickup, Mark, and Richard Johnston. 2008. "Campaign Trial Heats as Election Forecasts: Measurement Error and Bias in 2004 Presidential Campaign Polls." *International Journal of Forecasting* 24:272–84.
Rigdon, Steven E., Sheldon H. Jacobson, Wendy K. Tam Cho, Edward C. Sewell, and Christopher J. Rigdon. 2009. "A Bayesian Prediction Model for the US Presidential Election." *American Politics Research* 37:700–724.
Rothschild, David M. 2010. "Forecasting Elections: Comparing Prediction Markets, Polls, and Their Biases." *Public Opinion Quarterly* 73:895–916.
Shaw, Daron. 1999. "A Study of Presidential Campaign Event Effects from 1952 to 1992." *Journal of Politics* 61:387–422.
Sides, John. 2014. "Four Suggestions for Making Election Forecasts Better, and Better Known." *PS: Political Science and Politics* 47:339–41.
Sides, John, and Lynn Vavreck. 2013. *The Gamble: Choice and Chance in the 2012 Presidential Election*. Princeton, NJ: Princeton University Press.
Silver, Nate. 2008. "Should We Be Discounting Obama's Lead? The Next and Hopefully Last Big Change." Available at http://www.fivethirtyeight.com/2008/06/should-we-be-discounting-obamas-lead.html.
———. "Methodology." Available at http://fivethirtyeight.blogs.nytimes.com/methodology/.
Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. 2012. *Prediction Markets for Economic Forecasting*. Working Paper 18222. Cambridge, MA: National Bureau of Economic Research. Available at http://www.nber.org/papers/w18222.
Tanenbaum, Andrew. "Map Algorithm Explained." electoral-vote.com. Available at http://www.electoral-vote.com/evp2012/Info/map-algorithm.html.
Traugott, Michael W. 2005. "The Accuracy of the National Preelection Polls in the 2004 Presidential Election." *Public Opinion Quarterly* 69:642–54.
Tziralis, George, and Ilias Tatsiopoulos. 2012. "Prediction Markets: An Extended Literature Review." *Journal of Prediction Markets* 1:75–91.
van Dongen, Stefan. 2006. "Prior Specification in Bayesian Statistics: Three Cautionary Tales." *Journal of Theoretical Biology* 242:90–100.
Voss, D. Stephen, Andrew Gelman, and Gary King. 1995. "A Review: Preelection Survey Methodology: Details From Eight Polling Organizations, 1988 and 1992." *Public Opinion Quarterly* 59:98–132.
Walsh, Elias, Sarah Dolfin, and John DiNardo. 2009. "Lies, Damn Lies, and Pre-Election Polling." *American Economic Review* 99:316–22.
Wang, Samuel S.-H. 2013. "Origins of Presidential Poll Aggregation: A Perspective From 2004 to 2012." Available at https://web.math.princeton.edu/~sswang/Wang_Origins_of_Presidential_poll_aggregation_7oct2013.pdf.
Wilks, Samuel S., et al. 1948. "Report on the Analysis of Pre-Election Polls and Forecasts." *Public Opinion Quarterly* 12:599–622.
Wlezien, Christopher, and Robert S. Erikson. 2004. "The Fundamentals, the Polls, and the Presidential Vote." *PS: Political Science and Politics* 37:747–51.