Article:
Evans, J (2015) Evolving approaches to election forecasting. Italian Journal of Applied Statistics, 25 (2). pp. 105-121. ISSN: 1125-1964

EVOLVING APPROACHES TO ELECTION FORECASTING

Abstract. *Since the development of electoral forecasting as a formalised modelling process, rather than informed punditry, polling estimates and economic predictors have dominated a field developed principally around the U.S. case. The main approaches have favoured two-party/-candidate systems and looked at incumbent / opposition vote change as a zero-sum game, a focus which continues to guide much forecasting work. This paper considers the developments in election forecasting, both in the use of increasingly sophisticated modelling to arrive at more complete predictions of party vote shares beyond simple incumbent reward / penalisation approaches, and in the use of a wider variety of data sources in predicting outcomes. Whilst more demanding for accurate prediction, the development of such models should provide a better understanding of "what matters" in elections.*

Keywords: *elections, forecasts, economic models, polls*

1. INTRODUCTION

Among electoral specialists and political scientists, election forecasting is still a minority activity. Whilst election outcomes have been predicted as long as elections have been held, the development of a structured, rigorous approach to prediction rather than punditry – an individual's variably informed guess as to a likely outcome – has only developed more recently and been dominated by approaches trialled and refined on the United States. Indeed, the vast majority of election forecasting, and the work underpinning this, has until the last few years focused on the U.S. and systems with traditionally similar competitive structures, for example the United Kingdom. There has been a noticeable bias towards forecasting systems with a two-candidate or two-party format, or more precisely systems on which a two-party or two-bloc logic can be imposed. Whilst many of the assumptions connected to such choices are methodologically appropriate, equally the dominance of binary approaches has often led to a wider range of countries either being analysed in a way that

dampens the true complexities of a multi-party system or, because these complexities cannot be collapsed into such a format, have been overlooked altogether. Only lately have researchers interested in forecasting elections begun to explore other ways of forecasting election outcomes for such cases. A further development in forecasting has broadened the types of information used to build forecasts. For most of forecasting's history, political opinion polling constituted the main empirical source of prediction. Whilst polls still constitute the most accessible form of predictive benchmark – as pollsters go out of their way to emphasise, polls are not a forecast per se but rather a measure of the state of electoral play – other approaches, including econometric models, have complemented polls from the 1970s onwards. More recently, the accessibility of online surveys and social media data has provided a much wider set of sources from which to model elections predictively.

This article looks at the range of mainstream approaches to electoral forecasting to understand some of the advantages each offers and also the theoretical and methodological limits which render them appropriate for some forecasting contexts more than others. It will focus wherever possible on *ex ante* forecasting, given that in many respects *ex post* retrodiction resembles explanatory modelling rather than true prediction. It will underline some of the broader implications of assessing election forecasts to consider the means by which we may consider whether a forecast has been successful, accurate or informative – three terms which are by no means synonymous. It will also consider how some of the principal challenges of forecasting lie in adapting techniques designed for other purposes and which offer a potentially fruitful future for researchers looking to improve forecasting, and thereby add to our knowledge of how elections work. If there is one vital outcome from a useful election forecast, it is understanding how elections really are decided, both by the voters who participate and the systems in which they take place.

2. POLLING AND ELECTORAL FORECASTING

The use of polling for forecast models suffers from a retrospective asymmetry: incorrect polling forecasts are remembered long after correct forecasts are forgotten. One example of this is the now infamous photo of Harry S. Truman brandishing a copy of the Chicago Daily Tribune, emblazoned with the

headline “Dewey Defeats Truman” after his 1948 defeat of his Republican challenger, Thomas E. Dewey.

Yet, even in 1948, the use of polling was an attempt to bring rigour to election forecasting that local knowledge or informed but unreplicable punditry could not. In particular, the knowledge that voting was not an election-specific phenomenon, but one that could be systematised and therefore potentially predicted – “the existence of an extraordinarily uniform voting pattern”, as one of the forecasting pioneers had it (Bean, 1948: 6) – flagged the potential value of suitably designed polling. Whilst a similar sense of a system, and a baseline consistency to voting behaviour, would inform the large-scale survey approaches from the Michigan Center for Survey Research that went on to dominate explanatory models of voting,[1] the polling which had failed so badly in 1948, and previously, received a health-check from Frederick Mosteller and the Social Science Research Council team commissioned to review the poll failures (Mosteller *et al*., 1949). The resulting eight accuracy measures offered a series of post-election diagnostics as to the accuracy of polling based upon absolute and proportionate differences in polled and observed scores and were still widely used six decades later.

Most commonly, Mosteller 3 – the mean difference between candidates’ poll and election scores – and Mosteller 5 – the absolute difference in the two leading candidates’ poll and election scores – have been used to judge polling accuracy (Traugott, 2005: 645-648). As Mitofsky, who also favours the use of methods 3 and 5, and method 5 for multiparty as well as two-candidate races (1998: 241), shows, the advantages and disadvantages of the eight Mosteller measures depend on type of race, the treatment of undecided voters, and *ad hoc* decisions by the researcher, such as how many of the candidates to include for method 3, for example. Whilst very useful in their simplicity, the weaknesses of these measures includes their over-reliance on absolute scores and their generally favouring the analysis of predominantly two-party races: they often refer to ‘the two leading candidates’, giving no information about accuracy for other candidates, whether they account for 2% or 20% of the vote. Comparability of the measures across time and country is also *ad hoc*.

Until recently, however, a consistent benchmark applicable to any election has been hard to come by. Partly forecasters are saved by – or the victims of, depending on one’s perspective – the fickleness of political commentators and the electorate. As a means to an end, interest in election results decays rapidly

after the result is known. Incentives to hold pollsters to their results are often marginal in what can be a short-termist view, and polls need to have been significantly divergent from an election outcome for greater scrutiny to be brought to bear. In the long term, knowledge of pollsters' performance is vital, for an electorate to understand previous performance as a prior for the store placed in their estimates next time, and indeed for the pollsters themselves to identify consistently how they perform internally and against other pollsters.

Moreover, concerns over polling accuracy should be directed not so much at the direct accuracy of a poll as a forecast, but rather as a statistical estimate with associated error of a likely vote outcome in a known population. Martin, Traugott and Kennedy's implementation of an odds-ratio approach was the first – and relatively recent – measure to acknowledge this explicitly (Martin, Traugott and Kennedy, 2005). One of the key advances in polling accuracy has been the introduction of inferential statistical assumptions to polling samples. An intuitively appealing designation of a poll's respondents as a representative sample of the voting population, the log-odds transformation of a poll estimate and the eventual outcome provides a balanced index of polling accuracy.

In its application to date for multiparty systems (*e.g.*, Durand, 2008; Wright *et al.*, 2013) implementations of the A measure, generally termed A', remain essentially binary, contrasting odds of vote for one party with that of all others. Using an empty multinomial logit model, a full generalization of this approach for multi-party measures, leading to a summary index of overall poll accuracy, B, has been suggested more recently (Arzheimer and Evans, 2014). As a weighted or unweighted average of the absolute values of the *n* A' values, depending on how one wishes to treat the relative polling error for larger and smaller parties, B provides a similar type of estimate to Mosteller 3, but not dependent on absolute percentage scores, and one which takes into account the core of variation in any poll: sampling error.

Even allowing for sampling error, raw polls make naive predictors of election outcomes for a number of reasons. A range of exogenous variables introduce bias into polling estimates and reduce their validity. For across-time polls, changes in voter choice will of course change eventual accuracy. Whilst non-random sampling may always affect the accuracy, and in very close races sampling error may be enough to give a "false" outcome for general users of polls, systematic bias through house effects are sufficiently studied to allow an anticipation of how these may vary from a future election result. Where the

method applied by a polling house tends to overestimate a certain party's support, this may result from a methodological choice or in some cases a political choice. Whatever the reason, where polling houses adjust figures without an indication of the weighting applied or the unadjusted data, to allow the correction to be examined, the validity of their estimates is seriously compromised. As Durand has shown in the French case, polling estimates of the Front National Presidential candidate, Jean-Marie Le Pen, suffered across two elections with polling houses under-estimating his result in 2002, adjusting their figures for 2007 to offset a similar occurrence, and consequently over-estimating his performance (Durand, 2008: 283-284). Fortunately, in other systems methodological information is more freely available, and house effects are more stable and identifiable.

As a moving point estimate linked to election outcome, polling has remained the most widely used indicator in vote forecasting. Projecting vote from the historical association between trial-heat polls and Presidential election outcomes in the U.S. should provide a relatively accurate projection of future vote once the trial-heat at the appropriate time-point is known (Erikson and Wlezien, 2008: 194-195). Furthermore, in countries with established polling houses with clear methodologies, historical knowledge of these different houses' own biases and evidence of non-specific trends in vote intentions across time, poll estimates can provide an accurate forecast (*e.g.*, Fisher *et al.*, 2011).

The use of trial-heat polls as a forecast measure is the most direct use of polling in election forecasting. However, political polling covers a wider set of questions concerning candidates, parties and incumbent popularity, for instance. Whilst we would not expect that, for example, positive approvals for an incumbent constitute a proxy for vote, we should expect that this estimate of support is representative of a voting population and therefore constitutes a valid variable in the full calculus of vote. In particular, economic performance variables have constituted the main accompanying variables in this calculus.

3. ECONOMICS, MODELS AND EVALUATION

Political forecasting has taken its lead from evaluation and prediction in other disciplines, including public opinion and polling as a form of political market research, but also economics in the introduction of statistical models to identify

trends and extrapolate to future outcomes. The integration of the two into so-called "VP-function" (vote-popularity function) models moved election forecasting into the realm of statistical inference and away from simple numerical accuracy. Early work considered socio-economic and political factors that influenced either presidential poll ratings or the Congressional vote (Mueller, 1970; Kramer, 1971). Formalising the fluctuations in polling fortunes of a U.S. president demonstrated that poll ratings were not simply collective whims of the electorate, but rather systematic changes according to salient events and conditions at any given time. Similarly, the Congressional vote could be accounted for by incumbency and a series of economic variables.

Whilst some studies continued to look at popularity as a dependent variable (*e.g.*, Sanders, 1991), contemporary use of popularity has generally shifted this variable from the left- to the right-hand side of the equation for vote forecasting. Popularity is now seen as a useful complementary predictor of vote-share in a forthcoming election. The VP-function literature, to which econometric models owe their specifications, speaks in general of incumbent penalisation (or reward, if we set aside the asymmetry tenet), in favour of the opposition. Alternatively, aspects of economic development, such as variations in unemployment or per-capita GDP, are linked to opposition performance. The full range of economic indicators which have been developed in the explanatory VP-function literature – retrospective and prospective economic evaluations, sociotropic and egocentric evaluations of the economy, short-termism in voters' evaluations ("myopia") – find their way into forecast models (Dubois and Fauvelle-Aymar, 2004: 208-9). Voters punish incumbents for poor economic performance (but do not necessarily reward them for good performance). Voters' assessment of performance looks to the latter *within* the year of the campaign. Voters tend to look to the economy rather than their own finances as the key indicator. Such tenets have moved from explanatory to forecast models.

So, however, has the "apparent" and "inherent" instability described by Lewis-Beck and Paldam for these models (2000: 119-120).[2] Whilst the above knowns provide a broad basis to set up a forecast model, the exact specification is often less clearly defined. For example, September trial-heat polls and second-quarter GDP growth rate of the election year have been shown to provide an optimal forecast combination in accuracy and lead-time for the U.S. (Campbell, 2012). Clearly, for other countries, such a model may be inappropriate. First, clarity of economic responsibility needs to be established

(Bingham Powell and Whitten, 1993; Whitten and Palmer, 1999). Secondly, GDP growth may not theoretically constitute an important variable in the country or election in question. These issues may not be inherent to a well-specified model, however they are inherent to electoral contexts which change by period. Indeed, a well-specified forecast model for one election, let alone one country, may well be not be replicable elsewhere. Moreover, the statistical error which in an explanatory model may still be acceptable alongside significant effects from independent variables can be such that estimates from a forecast model become poor.

The differing priorities for model fit are reflected in the statistics used to evaluate forecast models. Alongside standard individual residuals in *ex post* predictions, accompanied by measures of in-sample error, mean absolute error (MAE) and out-of-sample measures look at forecasting success. Of greater interest, however, is how well models outperform other, "naive" predictors. In particular, the random walk and random mean models look at whether forecasts provide added value over simple lagged or regressive specifications. Forecasts are thereby judged not just in how well they estimate an outcome, but also by the value of the estimation process added to known prior information immanent from previous results.

In addition to such statistical diagnostics, a range of more qualitative indicators have been widely used in assessing the strengths of a particular model. Accuracy has evidently been a principal indicator of interest with regard to polling. However, as Lewis-Beck has noted (1985, 2005), there are three other criteria against which we should judge the quality of a forecast: lead-time, parsimony and reproducibility. Lead-time assesses the length of time before an election a forecast can be made. Parsimony looks at the number of predictors which the forecast needs to include to derive its estimate and "how much can be done with how little". Finally, reproducibility summarises clarity and simplicity,[3] and from that the capacity of a researcher to replicate the forecast approach both for the election in question and for future tests. Evidently, the Holy Grail of forecasting is a highly accurate forecast, estimated a significant period before the election using a small number of easily available predictors in a non-complex equation. Needless to say, such a unified equation has not been found, and researchers have spent a good deal of time in implementing trade-offs between the lead, parsimony and even reproducibility in looking to

maximise accuracy. However, as we shall see, even accuracy itself does not necessarily possess a fixed definition.

4. WHAT SHOULD WE FORECAST?

In contrast to explanatory models which are by definition retrospective, and therefore require no *ex post* evaluation,[4] true *ex ante* forecast models can and should be assessed for their efficacy once the result is known. At the basic level, a forecast which miscalls an election result is less useful than one which gets it right. Moreover, a model which even in *ex post* terms provides a highly accurate account of previous results loses much of its power at the first "wrong" *ex ante* result. In the case of small-*n* forecasts, for example legislative election results in most democracies, the threat of overfitting is ever present. Where n is only a small multiple of k, researchers need to be wary of investing too much predictive expectation, and particularly not be tempted to see the forecast fitting process as a way of discovering an as-yet latent relationship between predictors and vote.

Considerations of lead-time, in particular, occupy forecasters trying to identify the best set of predictors to include in a model. Where repeated tests indicate a predictive strength to the choice of a particular variable across time, its use is likely to continue even if the theoretical rationale for its use may become obscure. As we have seen, the use of second-quarter economic measures has such a rationale in the U.S., but in many other cases theoretical determinants of absolute or marginal levels of an economic variable, period of measurement, and even which economic variable to choose are often less clear. Theories also change across time. Issue ownership has historically led to inflationary indicators to be used for right-wing parties, and unemployment for those of the left. However, this delineation has become increasingly blurred, such that unemployment has now become a valence rather than positional indicator for both sides of the political spectrum, and in forecasting its relationship with individual parties a more complex one (Arzheimer and Evans, 2010: 25).

From a Friedmanesque forecasting perspective, the "why" of the good predictor is irrelevant, or at least secondary to its accuracy: it does, and that is what matters. From this perspective, a data-mining approach to finding the

clearest signal amidst noise could be justified. Again, however, with the paucity of cases which researchers are generally presented with in macro-level electoral data, signals may be misleading. A sample of *n* cases best fitted by a model with economic indicator X_t easily becomes a sample of *n*+1 cases best fitted with indicator $X_{t\text{-}6months}$. The expectation of a latent "true" relationship is unobservable, at best, and fallacious at worst. The use of meso-level data gets around the small-*n* problem, but complicates the estimation and adds to the noise.

At first sight, to ask what we are forecasting seems an odd question. But "the result of an election" is not a fixed measurement. Even in a straight Presidential run-off, two outcomes are relevant and equally important: the winner and the score. The results are mutually dependent, and may determine the function used for forecasting, but in terms of accuracy they are not the same. Whilst a win requires a score greater than the 50% cut-point, it is perfectly possible to have two competing models with one more accurate as to the eventual winner, the other more accurate as to the voting score. A simple example suffices:

	Model A	Model B	Outcome
Candidate 1	51%	44%	49%
Candidate 2	49%	56%	51%

Model A forecasts the outcome of the election incorrectly but is only 2 points out as to the final score. Model B forecasts the outcome correctly, but is over twice as inaccurate in vote share. Which model is better? To the lay person, more concerned by winners and losers, Model B gets the election right. However, there are many circumstances where Model A could claim some value, even if the prediction falls "on the wrong side of the line" (Jérôme and Jérôme-Speziari, 2012). A model which well ahead of time reveals information about a forthcoming election that other techniques did not bears consideration for use in subsequent elections. Benefits from lead-time accrue which are not necessarily related to accuracy. Campbell (2004: 733) is clear that the measure of accuracy depends on what the model is designed to do: if it is designed to predict vote share, rather than winner, then failure to do the latter if the former is satisfied is not grounds for rejecting the model. For our lay consumer of electoral forecasts this may seem odd, but the win/lose binary function at a 50%

cut-point is an artefact of both the electoral system and, somewhat ironically, a probabilistic function that only a minority of forecast models employ, linear regression still being the technique of choice for many forecasters.

This concern of "what are we forecasting?" can be broadened further. Adversarial two-party systems such as the U.S. have perhaps over-emphasised the importance of "who wins". Multi-party systems with coalition formation will almost certainly not be concerned with 50% + x as an indicator of victory, or only in exceptional circumstances. The temptation in forecasting multi-party elections is to make a simplifying assumption which converts the multi-party approach to a pseudo-two-party forecast, either functionally or in an underlying explanatory relationship. Most commonly forecasts of French elections, for example, have traditionally used the dominant bipolar characterisation of the French system to justify grouping parties into blocs of the left and right (Dubois and Fauvelle-Aymar, 2004; Foucault and Nadeau, 2012). In terms of extrapolating to the result of the final run-off, such a choice is entirely justifiable. However, it limits the forecast potential for a race which commonly includes at least ten candidates, and not all of them helpfully conflated into simply "left" or "right". Even in forecasts for many complex multi-party systems, models restrict themselves to one bloc's share of the vote or that of the incumbent party or coalition (*e.g.*, Lewis-Beck and Bélanger, 2012: 768) or, in one more innovative model, grand coalition partners in Austria (Aichholzer and Willmann, 2014). Both approaches are theoretically justifiable and internally consistent, so long as a bloc (rather than party or candidate) result is useful and externally consistent. Nevertheless, much of the information inherent in the election result is lost.

A winner may well be of importance in terms of highest vote share, but again this is not necessarily a given, especially in systems with pre-election coalition formation, and a strong but isolated party destined for opposition. Second-order elections such as the quinquennial European Parliament ballots derive little practical value from identifying winners, as opposed to vote and seat shares (Auberger, 2005). Under such circumstances, accuracy of vote share is paramount. However, whilst a global model would look to provide a measure for all parties, many models in fact only look to estimate that of one party, and not a party describable in any sense as a "winner". Instead, focus shifts to third parties, with the forecasting literature presenting a number of examples of

specific parties, of interest for reasons other than governing potential, being the subject of forecast modelling (Bélanger *et al*., 2010; Evans and Ivaldi, 2012).

In the case of work on forecasting far-right parties' scores, interest has lain in assessing the stability of votes for parties which in the past have been described as "flash" or "protest", with the inherent instability and apparently fluctuating electoral fortunes these labels suggest. The disconnect between much mainstream academic analysis of the far right – perhaps disproportionate to its electoral fortunes – and broader media commentary – cyclical in identifying "shock" performances – finds a helpful midpoint in forecast models' identification of a stable predictability to electoral performance based on traditional VP-function considerations. That the performance of such third parties does not necessarily impinge directly upon "winner" outcomes does not render their result any less integral to the overall election result.

In short, electoral forecasting does well to move on from single-equation, single-party/-candidate approaches to more complete models of overall electoral outcomes. This is not to deny the value of single-party estimates in predicting the winner. But forecasting can go beyond this one, obvious aim. Yet very few models look to estimate multiple outcomes through a single constrained model. Such approaches as Seemingly Unrelated Regression provide the methodological means to do precisely this. As Tomz *et al*. showed (2002), this stacked approach to model multiple party outcomes using the appropriate regressors, and accounting for by-definition correlated errors, is a relatively simple but appropriate solution. The challenge of specifying the relevant equations to multiple related outcomes is a more challenging task and perhaps accounts for the reason why such implementations to date in forecast models have been relatively rare (*e.g*., Arzheimer and Evans, 2010; Jérôme and Jérôme-Speziari, 2010). Nonetheless, they would seem vital to a global forecast.

Beyond the U.S., the use of modelling to provide more global accounts of outcome, including multiple party vote and seat shares, has developed most consistently perhaps in the recent 2015 U.K. General Election, with almost a dozen published forecast models, including a number of inter-institutional teams using adjusted polls, national- and constituency-level data and historical trend analysis to predict the outcome, both for "national" parties and for regional parties contesting only small subsets of the total territory (e.g., Fisher, 2015; Hanretty et al., 2015). Whilst for this election the underlying biases in

polling were a salutary lesson in the effects of error on vote estimation and, by extension, that of seats, the rigour and transparency apparent in these models provides a benchmark for future forecast endeavours.

5. THE GROWING CHOICE OF FORECASTING TOOLS

Perhaps the most overwhelming concern for *ex ante* forecasters is a property which overlaps lead-time, reproducibility and parsimony: simplicity. Whilst the researcher cannot evaluate accuracy ahead of an election, other than comparisons with exogenous benchmarks which themselves are forecasts, or with reference to immanent diagnostics that derive *ex post* from the model itself, the other three criteria can be evaluated or at least anticipated. To the extent that complexity of technique can be assumed to be cost-free – functional forms, estimation and the like for election data can probably now be assumed to be within the computational grasp of the majority of researchers, with little or no data retrieval cost beyond time – there is nevertheless still a cost to complexity involving greater quantities of data: parsimony. The use of ever larger numbers of variables – even should the degrees of freedom allow it – or increasingly complex assumptions, transformations and the like, brings the inherent risk of being irreproducible in the future, or reproducible only with difficulty in the present data by other researchers. A simple model does not eliminate that risk entirely, but replicating the model exactly in the future is less prone to error or confusion. Rather than necessarily pursuing grander, more sophisticated models doing the same thing as simpler models, an examination of alternative approaches, both in terms of method and sources of data, is a useful complement.

In terms of models approaching the criteria set out by Lewis-Beck, there is an absence of what might be deemed "qualitative" approaches, or forecasts which do not base themselves on a numerical score estimated mathematically. There is one notable exception, namely the "Keys to the White House" model developed by Lichtman (2008), which identifies 13 true/false "keys" of which six or more need to be turned to false for the incumbent party to lose. The keys themselves are statements regarding policy performance, social unrest, campaign and candidate personality characteristics, each with a binary outcome. This model closely meets Lewis-Beck's criteria: the model is accurate, having correctly predicted all popular vote share majorities from 1984

to 2012.[5] It is parsimonious in that all 13 keys have been necessary in at least one election – although from a statistical point of view, a case might perhaps be made that the loss of certain keys would reduce fit but leave a nonetheless strong forecast model. Lead-time may be extensive, given keys may be turned well in advance of the election. Reproducibility is also strong, although critics have noted the subjectivity of some of the keys, for example candidate charisma.

The U.S. advantage is clear here too. We are unaware to date of any similar algorithmic approach to forecasting in other countries. The stability of the keys in the U.S. case has been demonstrated in a stable competitive system since 1866 across 39 observations. The outcome of the election, occasional third-party candidates notwithstanding, remains firmly dichotomous. The forecast gives no indication of popular vote share; within its own terms, it indicates who will receive majority support.

Paradoxically, perhaps, in the quest for a stable forecast equation, using scientific approaches to forecasting, one of the more successful areas complementing popularity and vote-intention polls and economic predictors has been what is termed the "wisdom of crowds" approaches. Rather than polling individual voting intentions, a sample of the electorate are asked to estimate the outcome of the election, either nationally or locally (Lewis-Beck and Skalaban, 1989; Murr, 2011). Overall, individual biases – ideological, informational and otherwise – tend to aggregate out, resulting in generally accurate forecasts of outcome. When voters are asked what will happen, rather than what they will do, "bets beat polls" (Hofstee and Schaapman, 1990). "Expert" crowds have also been used as an aggregating method of vote forecast, with sometimes variable results (Sjöberg, 2008; Evans and Ivaldi, 2012: 60-62). In the latter case, whilst not as accurate as the econometric forecast, the expert forecast still fell well within 1 percentage point of the final outcome. As both Sjöberg, as well as Lewis-Beck and Skalaban found, however, there is little evidence that informed "insiders" perform any better than the general public.

The wisdom of crowds has also motivated the pioneering of political markets, either political betting or prediction markets, as a forecast tool. From the original Iowa Electronic Markets, individuals adopting a market perspective on purchasing political stock give an accurate aggregate account of not only the likely winner, but – where it has been trialled on multi-party systems – also relative support for other parties in the system. Technology has opened such

markets to worldwide participation, perhaps increasing less informed participants, but there seems to be little evidence of a reduction in forecast capacity from this.

In advances in forecasting, the intersection between technology and knowledge has opened other avenues for prediction. As technology blurs the lines of who may be classified an "expert" and who simply a member of the general public, so forecasting has exploited the rise of social media as a location of political commentary to examine its potential as a data source. The use of Twitter and Facebook has been mostly commonly found in analyses of party elite campaigning. However, through the use of sentiment analysis, looking at content and colour of political comment in social media, relatively accurate estimates of subsequent election outcome appear possible (Franch, 2011; Ceron *et al*., 2014). Political social media users in no way constitute a representative sample of the electorate, but the proportions of positive and negative comment across the spectrum to date appear to line up closely with eventual position in the race.

Having begun our discussion with the original naive forecast – the univariate opinion poll – the most involved and, by many measures, the most successful forecast technique implements a crowd logic to forecasting by issuing an aggregated prediction based upon a range of individual forecasting approaches. The most famous forecasts using this approach, Pollyvote, combines four forecast approaches – trial-heal polls, expert judgement, econometric model forecasts, and political markets – to issue a combined forecast (Graefe *et al*., 2014). The logic is similar to the wisdom of crowds, in that sources of error and bias specific to each poll will largely cancel out through aggregation. The overall success of the approach is revealing through one methodological choice: all component are equally weighted (2014: 46-47). With no *a priori* information to suggest favouring any one forecast approach over others in the final prediction, the information and error of each is given equal weighting.

Indirectly, this decision illustrates the inherent instability to date in all individual forecast approaches. Across time, no one model has managed to predict election outcomes either within or across countries better than all other approaches. Had it done so, projects such as Pollyvote would not include four "schools" amongst its components, and the forecasting literature would be less broad, and have an even more minority status than it currently does. Once

again, the aggregating approach is a step forward for the U.S. case.[6] With the diversity of approaches available to the potential forecaster, other country cases would benefit greatly from similar methodological innovations and advance.

6. CONCLUSION

As a minority activity, election forecasting remains in many ways quite a conservative domain. Having identified a set of imperfect but broadly consistent predictors in the economy, incumbency and polling estimates, the majority of forecasters continue to use these as a baseline model, even if they may introduce additional variables or methodological tweaks to the functions. Many of the innovations highlighted above remain a relatively esoteric preserve for a few researchers. Whilst knowledge of individual voters' decisions in the voting booth will never be error-free, further developments in analysis of election context give us variables which have yet to penetrate the mainstream forecasting literature. A small but highly consistent literature on the effects of candidate attractiveness, for example, is still generally the preserve of the *ex post* explanatory model (Rosar *et al*., 2008; Mattes and Milazzo, 2014). Lead-time may be limited by candidate selection processes, but potentially these insights could be used more commonly in an *ex ante* forecast.

In the introduction, we specifically indicated that *ex ante* forecasts were to be favoured over exclusively *ex post* retrocasts. Whilst the acid test of a successful forecast model is its ability to predict an election in the future, we should nonetheless not be misled by the absence of a known result into regarding such forecasts as free from specification adjustment and other data-fitting, which *ex post* forecasts can never entirely disprove. Election specialists intent on forecasting an election will have a host of information at hand which will guide the specification of a model and the evaluation of its projected outcome. Unexpected results are likely to be revisited rather than released.

The only guard against such anticipatory adjustment is replication of the identical model on subsequent elections. In their overview of the VP-function literature, Nannestad and Paldam highlight this as a similar concern for comparative explanatory models: "Econometrics is so flexible as to allow the

hardworking researcher to find something if he mines the data hard enough by experimenting with the lags, the periodization, alternative series, special events dummies, etc. In short: those who seek shall find […] In order to be convincing, the functions should be exactly the same; but then countries are different, so there is a real dilemma" (1994: 234-5). As well as countries being different, so are elections within countries. Yet, once again, the infrequency of elections and changes in the competitive properties of these races makes such replication extremely difficult, and for the purposes of good forecasting – if not good forecast models – the opposite motivations apply.

Going forward, then, the continued instability of specific models appears to be a given which requires adjustment by time and place. It is unlikely that, with intelligent *ad hoc* adjustment, such models will fail to provide reasonable estimates of winners, incumbent performance and vote share. A useful complement to this in our view would be a shift to considering more complete accounts of election performance across the "neglected democracies" (Lewis-Beck and Bélanger, 2012) but also in countries where forecasting has become more mainstream, if still minority. Identifying the key empirical predictors across a broader range of systems, and parties, would provide significant advances in our understanding of the mechanics of elections.

NOTES

1. And incidentally, whose founding father Angus Campbell managed to predict the Dewey-Truman outcome correctly (Campbell and Kahn, 1952)
2. More recent work by Bellucci and Lewis-Beck (2011) argues that this instability can be substantially reduced through improvements in data and model specification.
3. In fact, in the original derivation of the evaluation criteria, Lewis-Beck uses clarity, usability and specification separately, instead of reproducibility (Lewis-Beck, 1985: 60). Given that we do not use the possible formula for combining the criteria here, the conflation into general reproducibility has no substantive impact on the discussion which follows.
4. We refer here to an assessment of accuracy, rather than standard model fit diagnostics.

5. As with the vast majority of all forecasts, the Presidential outcome *per se* evaded it in 2000.
6. The Pollyvote method has also been used on recent German national elections (Graefe, 2015).

REFERENCES

Aichholzer, J. and Willmann, J. (2014). Forecasting Austrian national elections: the grand coalition model. *International Journal of Forecasting* 30(1): 55-64.

Arzheimer, K. and Evans, J. (2010). Bread and butter à la française: Multiparty forecasts of the French legislative vote (1981–2007). *International Journal of Forecasting* 26(1), 19-31.

Arzheimer, K. and Evans, J. (2014). A new multinomial accuracy measure for polling bias. *Political Analysis*, 22(1), 31-44.

Auberger, A. (2005). Forecasts of the 2004 French European election. *Swiss Political Science Review* 11: 61-78.

Bean, L. (1948). *How To Predict Elections*. New York, Alfred Knopf.

Bélanger, E., Nadeau, R. and Lewis-Beck, M. (2010). Forecasting the vote for a third party: the British Liberals, 1974-2005. *British Journal of Politics and International Relations* 12(4): 634-643.

Bellucci, P., & Lewis-Beck, M. (2011). A stable popularity function? Cross-national analysis. European Journal of Political Research, 50(2): 190-211.

Campbell, A. and Kahn, R. (1952). *The People Elect a President*. University of Michigan, Survey Research Center.

Campbell, J. (2004). Introduction – the 2004 Presidential election forecasts. *PS: Political Science and Politics* 37(4): 733-35.

Campbell, J. (2008). Evaluating U.S. presidential election forecasts and forecasting equations. *International Journal of Forecasting* 24: 259-271

Campbell, J. (2012). Forecasting the presidential and congressional elections of 2012: The trial-heat and the seats-in-trouble models. *PS: Political Science and Politics* 45(4): 630-34.

Ceron, A., Curini, L., Iacus, S. and Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media and Society* doi: 10.1177/1461444813480466.

Dubois, É. and Fauvelle-Aymar, C. (2004). Vote functions in France and the 2002 election forecast. In M.S. Lewis-Beck, ed. *The French Voter: Before and After the 2002 Elections* Basingstoke, Palgrave Macmillan, 205–230.

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.

Erikson, R. and Wlezien, C. (2008). Are political markets really superior to polls as election predictors? *Public Opinion Quarterly* 72: 190-215

Evans, J. and Ivaldi, G. (2012). Forecasting the FN Presidential vote in 2012. *French Politics* 10(1): 44-67.

Fisher, S. (2015) 'Predictable and unpredictable changes in party support: a method for long-range daily election forecasting from opinion polls', *Journal of Elections, Public Opinion and Parties* 25(2): 137-158.

Fisher, S., Ford, R., Jennings, W., Pickup, M. and Wlezien, C. (2011). From polls to votes to seats: forecasting the 2010 British general election. *Electoral Studies* 30(2): 250-57.

Foucault, M. and Nadeau, R. (2012). Forecasting the 2012 French Presidential election. *PS: Political Science and Politics* 45(2): 218-222.

Franch, F. (2011) (Wisdom of the Crowds)^2: 2010 UK election prediction with social media. *Journal of Information Technology & Politics* 10(1): 57-71.

Graefe, A. (2015) 'German Election Forecasting: Comparing and Combining Methods for 2013. *German Politics* 24(2): 195-204.

Graefe, A., Armstrong, J. S., Jones, R. and Cuzan, A. (2014). Combining forecasts: an application to elections. *International Journal of Forecasting* 30(1): 43-54.

Hanretty, C., Lauderdale, B. and Vivyan, N. (2015, forthcoming) 'Combining national and constituency polling for forecasting', *Electoral Studies*.

Hofstee, W. and H. Schaapman (1990). Bets beat polls: averaged predictions of election outcomes. *Acta Politica* 25(3): 194-207.

Jérôme B., Jérôme, V. and Lewis-Beck, M. (1999). Polls fail in France: forecasts of the 1997 legislative election. *International Journal of Forecasting* 15: 163–174.

Jérôme, B. and Jérôme-Speziari, V. (2010). Forecasting partisan dynamics in Europe. *International Journal of Forecasting* 26(1): 98–115.

Jérôme, B. and Jérôme-Speziari, V. (2012). Forecasting the 2012 Presidential election: lessons from a region-by-region political economy model. *French Politics* 10(4): 373-377.

Kramer, G. (1971). Short-term fluctuations in US voting behaviour, 1896-1964. *American Political Science Review* 65: 131-143.

Lewis-Beck, M. (1985). Election forecasts in 1984: how accurate were they? *PS* 18(1): 53-62.

Lewis-Beck, M. (2005). Election forecasting: principles and practice. *British Journal of Politics and International Relations* 7(2): 145-164.

Lewis-Beck, M. and Bélanger, E. (2012). Election forecasting in neglected democracies: an introduction. *International Journal of Forecasting* 28(4): 767-68.

Lewis-Beck, M. and Paldam, M. (2000). Economic voting: an introduction. *Electoral Studies* 19(2): 113-121.

Lewis-Beck, M. and Skalaban, A. (1989). Citizen forecasting: can voters see into the future? *British Journal of Political Science* 19(1): 146-153.

Lewis-Beck, M. and Tien, C. (2005). The jobs model forecast: well done in 2004. *PS: Political Science and Politics* 38(1): 27–29.

Lichtman, A. (2008) *The Keys to the White House*. Lanham, Rowman and Littlefield.

Martin, E., Traugott, M. and Kennedy, C. (2005). A review and proposal for a new measure of poll accuracy. *Public Opinion Quarterly* 69(3): 342-369.

Mattes, K. and Milazzo, C. (2014). Pretty faces, marginal races: predicting election outcomes using trait assessments of British parliamentary candidates. *Electoral Studies* 34: 177-189.

Mitofsky, W. (1998). Was 1996 a worse year for polls than 1948? *Public Opinion Quarterly* 62(2): 230-49.

Mosteller, F., Herbert, H., McCarthy, P., Marks, E., & Truman, D. (1949). *The pre-election polls of 1948: Report to the committee on analysis of pre-election polls and forecasts*. New York, Social Science Research Council.

Mueller, J. (1970). Presidential popularity from Truman to Johnson. *American Political Science Review* 64: 18-34.

Murr, A. (2011). 'Wisdom of crowds?' A decentralised election forecasting model that uses citizens' expectations. *Electoral Studies* 30(4): 771-783.

Nannestad, P. and Paldam, M. (1994). The VP-function - a survey of the literature on vote and popularity functions after 25 years. *Public Choice* 79: 213-45.

Powell, G. Bingham and Whitten, G. (1993). A cross-national analysis of economic voting: taking account of the political context. *American Journal of Political Science* 37: 391-414.

Rosar, U., Klein, M. and Beckers, T. (2008). The frogpond beauty contest: physical attractiveness and electoral success at the North Rhine-Westphalia state election of 2005. *European Journal of Political Research* 47(1): 64-79.

Sanders, D. (1991). Government popularity and the next General Election. *Political Quarterly* 62(2): 235-261.

Sjöberg, L. (2008). Are all crowds equally wise? A comparison of political election forecasts by experts and the public. *Journal of Forecasting* 28(1): 1-28.

Tomz, M., Tucker, J. and Wittenberg, J. (2002). An easy and accurate regression model for multiparty electoral data. *Political Analysis* 10(1): 66–83.

Traugott, M. (2005). The accuracy of the national preelection polls in the 2004 presidential elections. *Public Opinion Quarterly* 69(5): 642-654.

Whitten, G. and Palmer, H. (1999). Cross-national analyses of economic voting. *Electoral Studies* 18: 49-67.

Wright, M., Farrar, D. and Russell, D. (2013). Polling accuracy in a multiparty election. *International Journal of Public Opinion Research* doi: 10.1093/ijpor/edt009.