BJPIR: 2005 VOL 7, 145–164

Election Forecasting: Principles and Practice

Michael S. Lewis-Beck

To forecast an election means to declare the outcome before it happens. Scientific approaches to election forecasting include polls, political stock markets and statistical models. I review these approaches, with an emphasis on the last, since it offers more lead time. Consideration is given to the history and politics of statistical forecasting models of elections. Rules for evaluating such models are offered. Examples of actual models come from the United States, France and the United Kingdom, where this work is rather new. Compared to other approaches, statistical modelling seems a promising method for forecasting elections.

> Humans are not very good at predicting the future, even just the next five minutes. (Gudmund Iversen)
>
> If you must forecast, forecast often. (Paul Samuelson)
>
> Prediction is central in science and in evaluating alternative generalizations or models. (Arnold Zellner)
>
> There are things we know we know. We also know there are known unknowns. (Donald Rumsfeld)

Forecasting aims to tell of events before they happen. It differs from *prediction* in that it looks to the future, whereas prediction may not (as in a successful reconstruction of some past outcome). Further, forecasting differs from *explanation*, having the goal of predicting an outcome, rather than the goal of theorising about outcomes. There are two types of forecasting, *scientific* and *non-scientific*. The latter are guesses based on hunches, intuition, stargazing, casual conversation, non-systematic interviews, insider talk or coincidence. The former, which concern us here, offer estimates based on some scientific procedure, such as a simulation, a market analysis, a sample survey or a statistical model. The major task at hand is explication of statistical models for forecasting elections. But, before embarking, it is worth considering the value of the enterprise.

Election forecasting serves many purposes, some public, some academic. It provides information to leaders and followers about a likely outcome, allowing them to make adjustments they feel necessary. With respect to political activists in particular, it might signal a redirection of their policy targets. For any interested party, tracking the election forecasts can assist them in evaluation of a campaign. In the community of scholars, these statistical models encourage careful measurement and theory building, including updating from the forecasting results. The aggregate unit—usually the nation—is the unit of analysis for the dependent variable, rather than the individual unit—the voter. Hence, much is learned about what causes election outcomes in democratic polities as a whole. Finally, good election forecasting models satisfy our intellectual curiosity—most of us, as citizens, want to know who is going to win and by how much. It is a piece of information that is intrinsically interesting in a healthy democracy.

Forecasts for a future election are always uncertain. For example, we do not know for sure, in advance, what the vote share will be for a governing party. In other words, our models are stochastic, rather than deterministic. Even if a model is correctly specified, there will be some randomness in the error term and in the parameter estimates. A further source of error comes from the measurement of the variables, especially those from poll data, with their built-in sampling error. Error is expected; the forecaster's goal is to minimise it. With respect to the general classes of forecasting models, some promise less error than others.

We can divide models into *conditional* versus *unconditional*, and *after-the-fact* versus *before-the-fact*. For a conditional model, the value of at least one of the independent variables is unknown, and so must be estimated, which increases error. For an unconditional model, the values of the independent variables are known (at least as observed), thus removing that source of error. With an after-the-fact model, all the observations, on the dependent variable as well as the independent variables, are known, thus rendering it unconditional. Such after-the-fact models can be useful for testing, e.g. assessing out-of-sample predictions of excluded observations in a cross-validation strategy. To illustrate, we could estimate a UK general elections model on 1955–1997 data, predict the now known 2001 result and assess the error in the point estimate.

While informative, after-the-fact tests do not forecast the object of ultimate interest—an election yet to occur. That is the task of a before-the-fact model. With such a model, the final observation on Y is unknown, while the observations on the Xs may be known or not. Here we focus on before-the-fact models with known (unconditional) Xs, because of the error reduction that affords, and because they have lead time (the final Y is not yet known). Thus, for the UK case under consideration, we hold relevant only models where all the independent variables have scores obtainable before the 2005 election takes place.

Developing a model is the first part of the process, and forecasting is the second. We take first things first. There are many kinds of possible models, from simple extrapolation of a time series to ARIMA models, from single-equation regressions to systems of simultaneous equations. We focus on single-equation regression models. For one reason, they are informed by theory. Other things being equal, a model based on sound electoral theory will perform better than an empirically-induced, atheoretical model. For another reason, the number of observations is small, precluding sophisticated time-series work or elaborately specified equations. In the case of the UK, for example, the available election series will likely have a count below 20 for years to come.

The steps for constructing a forecasting model are much the same as for any social scientific regression model—consult theory, express the theory in an equation, gather good data on the variables, estimate the equation, evaluate the estimates, judge the fit. However, because the goal is forecasting, the emphasis tends to differ. The focus is on Y, not on X. It is important to account for the dependent variable as fully as possible. This contrasts with the sometime-emphasis on the effects of a single X (in the face of controls). Special attention must be given, then, to Y and the question of model fit.

Assuming the model gives a good picture of the real world, the analyst can go on to forecast. In particular, the future value of Y can be predicted from the known values of X, with the reasonable expectation that the prediction will be an improvement, perhaps a great improvement, over an uninformed guess. How to evaluate the forecast? Obviously, the first thing to do is compare the estimate for Y to the observed Y. If the forecast is Labour with 40.5 per cent of the vote, and Labour actually gets 40.5 per cent, that is an accurate—indeed dead-accurate—forecast. But things are never that easy, or so easily finished. Accuracy is but one component, and accuracy can be defined in various ways. Moreover, there are components besides accuracy that go into evaluating the quality of a forecast, and the model that generates it.

Below, I review the leading scientific approaches to election forecasting. Then, I discuss the politics and history of the enterprise. Special attention is given to the statistical modelling of election forecasts. The principles of election forecasting are carefully developed, including the components of forecast evaluation and model revision. I eventually turn to the practice of election forecasting in the US and France. This provides context for evaluation of election forecasting in the UK, with applications to the upcoming 2005 general election contest.

Comparing Statistical Models to Alternative Approaches

After an election, statistical modellers are asked, 'How did you do?'. The response should be, 'Compared to what?'. For example, a certain statistical model that performs well compared to other statistical models, and also well compared to other approaches, may be judged excellent. However, another statistical model that is bested by most other models and opinion polls cannot be judged a success. There are various possibilities, so it is best to explore the terrain systematically. Besides the statistical modelling approach, which we will elaborate, three other scientific forecasting approaches merit mention here: vote intention surveys, vote expectation surveys and political stock markets.

In *vote intention surveys*, a sample of voters is asked something like: 'If the election were held tomorrow, which candidate would you vote for?'. The percentage vote shares in the responses are taken to forecast the final vote shares. Gallup is the classic example here, although now there are many other firms asking the same sort of question. In *vote expectation surveys*, a sample of voters is asked something like: 'Regardless of your personal preferences, who do you think will win the upcoming election?'. The distribution of responses serves to forecast the victor in the contest. The American National Election Study asks this question as do some commercial polling operations. In a *political stock market*, a group of traders invests money in candidates running for office. More money is invested in candidates the traders judge as more likely to win. A candidate's investment share provides a vote forecast for the candidate. The Iowa Electronic Market is the leading example of this sort of forecasting.

Each of these approaches has been applied to forecast specific elections. They do not always give consistent answers. Note that, in principle, each is susceptible to rules of evaluation, as will be discussed below. Note, further, that these three approaches—vote intentions, vote expectations, the political stock market—are not based on any theory of the vote. Instead, they are merely providing point estimates on a dependent variable. The statistical modelling approach, in contrast, relies on electoral theory for its construction. It is my belief that, in the long run, the statistical modelling approach, because its draws on voting theory, will yield a better performance. Time will provide us the evidence on that speculation.

The Politics of Election Forecasting

Election forecasting has a political side. Politicians may seek out your advice. As a scientific forecaster, it is important to convey a non-partisan image. If your work becomes identified with a particular candidate or cause, your credibility may be lessened. The media will also seek you out. That is their job, and as a forecaster it is in some sense your job to get your forecasts out before the election. In dealing with the media, it pays to take time to explain your method to responsive reporters. To the extent that they understand it, they will do a better story. For example, one thing that is difficult to convey is the probabilistic notion of a forecast. A standard error around a forecast, important as it is, is hard to grasp for most folks.

The idea of a unique point forecast presents a communication problem as well. Statistical models provide one point estimate. The measures, available at a certain time, are plugged into the model, and out comes the forecast. But, reporters come back repeatedly for updates. There are no updates, unless the forecaster had used a preliminary estimate for, say, GNP, or is now using a more current estimate. Of course, these slight measurement revisions, and subsequent forecast revisions, are a source of potential confusion to the media. Moreover, that confusion is compounded by the fact that pollsters now provide new survey results on an almost daily basis. They come to expect modellers to do the same.

As a forecaster, you must be aware that the media may 'use' you. Your forecast is 'news', and as such they may wish to manage it a bit. Sometimes my forecasts have made a front-page splash. At other times they have actually been suppressed, the worst thing because it deprives you of an expected, and perhaps the only, publication outlet. Once, in France, a leading conservative daily declined to publish the article they had commissioned from me, because they did not like its forecast of a Socialist win. A comparable incident happened to me in Hungary. Perhaps these incidents stem from fear that the forecast will influence votes. My view is that my forecasts, on balance, have no effect on the outcome of an election. That is mainly because there are many competing forecasts, leaving the average voter with neither the time, nor the desire or the ability to sort them.

Further, the average voter sometimes simply disregards forecasts because they believe the forecaster is intentionally biased in favour of a particular candidate or party. It is not uncommon, when I give public talks, for someone in the audience to challenge me on why I am for the candidate I forecast. It takes work to explain to them that the forecast comes out of the calculations in the model, not out of my own personal political calculations.

Still, there are voters who believe that a forecaster is a powerful political activist. I have received a surprising amount of 'hate mail', nasty phone calls or hostile emails, attacking me personally in the strongest possible language for my forecasts. Here is an example of one, not the worst (the typos are in the original):

> Dear Sir,
>
> I heard you on the Tom Clark program (the ultimate leftist politically correct lair) on WPR. It is shocking to see how Universities are abandoning scholarship for social engineering of their students. One has to wonder how the Democratic party will express its thanks to the U of I ... for you book predicting Gore as the coming winner of the presidential election.

In sum, election forecasters place themselves in the political arena, and must be prepared for the play of the game.

The History of Election Forecasting

In western democracies, guessing who will win elections is a long-standing tradition. However, scientific efforts are rather recent. Beginning in the 1940s, economist Louis Bean (1948) tried to predict US elections by identifying bellwethers, in particular at the state level. A bellwether state of course cannot provide a true forecast, since its performance registers itself only after the election. With respect to before-the-fact forecasts, the Gallup Poll first released a presidential pre-election survey in 1936 (its final survey missing the Roosevelt vote share by 7 per cent). After the war it made its now infamous prediction of a Dewey victory in 1948. Thus, its track record as a reliable guide to the winner, if not the winner's precise vote share, did not come to be established until the end of the 1950s (see *Gallup Reports*).

Statistical forecasting models started to appear around 1980, in the works of Sigelman (1979), Lewis-Beck and Rice (1982), Rosenstone (1983) and Lewis-Beck and Rice (1984). There are now many such models for US elections, and much of that literature is summarised in books by Campbell and Garand (2000) and Lewis-Beck and Rice (1992). In the late 1980s, a 'citizen's forecasting' approach was developed, focusing on pre-election presidential vote expectations among samples of the electorate (Lewis-Beck and Skalaban 1989; see also Lewis-Beck and Tien 1999). At about the same time, the stock market notion of election forecasting was being established (Forsythe et al. 1989).

Scientific election forecasting has of course taken place outside the US, although this has been almost entirely based on the use of pre-election polls. Today in virtually every democracy Gallup, or a rival commercial firm, conducts and publishes public opinion surveys of vote intentions prior to a major election. For example, in the UK, such pre-election polls began as early as 1947. By way of contrast, statistical forecasting models in these countries are scarce; only for France and the UK are there more than a few isolated studies. With respect to other approaches, such as vote expectation models or political stock markets, there is virtually nothing. One novel approach—neural networks—has had a unique trial in the UK, and is the subject of one of the articles in this collection.

Below, we review the core specification behind the statistical modelling approach, and go over relevant measurement and data issues. We begin with the US case, where there has been the most work. After a thorough evaluation of US presidential election models, we compare their performance to alternative approaches. Then, we move to the French case and, finally, the UK case. For each, the evolution of these models has been somewhat different.

Model Specification, Measures and Data: The US Example

Most models have at their core the assumption that Vote = f(politics, economics). Verbally, the typical equation reads something like:

$$\text{Government Support} = \text{Government Performance} + \text{Economic Performance} + \text{error}. \quad \text{(Eq.1)}$$

While this may appear a version of 'election-as-a-referendum-on-the-government' theory, popularised by Edward Tufte (1978), it is compatible with other theoretical approaches, especially as the particular variables chosen change. Most often the variables are something like:

$$\text{Incumbent Vote} = \text{Incumbent Popularity (lagged)} + \text{Economic Growth (lagged)} + \text{error}. \quad \text{(Eq.2)}$$

An example from the US, for presidential elections, is as follows (Lewis-Beck and Tien 2000, 84–86):

$$V = 37.31^{*} + 0.28^{*}P + 1.32^{*}E + e \quad \text{(Eq.3)}$$

$$\text{R-sq.} = 0.84,\ \text{adj. R-sq.} = 0.79,\ \text{SEE} = 2.82,\ N = 11\ (1948\text{–}1988)$$

Where V = presidential party share of the two-major-party popular vote in the November election, P = presidential popularity, or percentage who approve of how the president is handling his job in the July Gallup Poll of the election year; E = percentage growth in real GNP over the first two quarters of the election year; e = error; * = statistically significant at 0.05.

There is considerable, although not complete, agreement about the core specification. This core 'political economy' model may have added to it institutional variables, usually something to do with incumbency advantage (disadvantage) or term limits (Lewis-Beck and Tien 2002).[1] Further, some authors look essentially at economic determinants (Fair 1978; Hibbs 1982), whereas others just at political ones (Brody and Sigelman 1983; Norpoth 2004).

Also, the economic measures differ: GDP growth (Abramowitz 2004; Campbell 2004b); GNP growth (Lewis-Beck and Tien 2004); perception of personal finances

(Holbrook 2004); prospective personal finances (Lockerbie 2004); leading economic indicators (Wlezien and Erikson 2004); income growth (ibid.); job growth (Lewis-Beck and Tien 2004). There is more agreement on political measures, which usually tap presidential popularity. However, there is disagreement over the lag structure of popularity, not to mention over the lag structure of the economic variables.

Finally, the models are not always estimated on the same data base. Most use data from 1948 to the present. But, some start as late as 1956 (Holbrook 2004; Lockerbie 2004), some as early as 1912 (Norpoth 2004). This means not only different sample spaces, but sometimes samples which are double in size, from one study to another. These data differences, taken together with the difference in measurement, and disagreements over specification outside the core model, can produce forecasts which differ widely.

Forecast Evaluation

A forecasting instrument should be evaluated by four criteria: *accuracy, lead, parsimony* and *reproducibility* (Lewis-Beck 1985a; Lewis-Beck and Rice 1992, ch. 6). In forecasting, a necessary evaluation standard is accuracy. If a model routinely produces inaccurate forecasts, it is not good. Election forecasts that are many points off are of little help. Below, we eventually review in detail the criteria for this paramount condition. However, accuracy is not a sufficient condition. Besides accuracy, a model must have lead, i.e. the forecast must be made before the event. The farther in advance a model produces accurate forecasts, the better.

With respect to elections, however, there are limits to lead time. One fixed legal limit is the official calendar distance between elections, e.g. five years. A practical limit is the relevance of the X scores many years from the election itself. Forecasts from three years out, for instance, would be vulnerable to charges of dangerous extrapolation. Still, giving the forecast a full horizon, say six months to a year, allows for an impressive performance (in part because assumptions about campaign equilibrium over the unmeasured final period become plausible). Further, forecasts with considerable lead help solve the endogeneity bias in parliamentary systems, i.e. the problem of a government calling an election with an eye to the X scores. When the lead time is short, say a month or less, the forecasting exercise itself risks being trivial. Little is gained from a model that forecasts a few days before the election. Such models may be theoretically tautological, not to say empty, and lack the anticipatory feel expected from a true forecast.

A forecasting model should also be parsimonious, as the principle of Ockham's razor reminds us. Other things being equal, a few well-specified variables will work better than many questionable ones. The parsimony admonition bears on the classic regression assumption of no specification error. If a model restricts itself to independent variables based on strong theory, it should maximise its performance, especially assuming those variables are not correlated with the error term. Support for this argument comes from recent evaluations of large-scale econometric models which, along with certain high-tech advances, have not significantly improved macro-economic forecasting in 50 years (Armstrong 2001; Holly and Weale 2000).

In addition, fewer parameters to estimate mean less error from parameter estimation itself. As a practical matter, parsimony is necessary with these election forecasting models, given that the sample sizes are so small (<20). With more than a handful of predictor variables, the degrees of freedom are quickly exhausted and forecasting becomes untenable.

The standard of parsimony bears on our last standard, that of reproducibility. In general, parsimonious models are easier to understand, and therefore easier to reproduce. Any election forecasting model needs to be reproducible, by the author and other analysts. This is less likely to happen if the measures are costly, in time or money, for example, if they are rather obscure indices that resist reconstruction or replication, or literally require many pounds sterling to purchase. If measures are difficult to obtain, then the model's reproducibility, and hence its scientific integrity, are at risk. Finally, it goes without saying that the measures must be reproducible *before* the event forecast. Fortunately, many current forecasters employ data that are readily available in academic or government archives well before the elections take place.

The criterion of accuracy we have saved for last, because it deserves special attention. A model that fits better, forecasts better. In classical regression analysis the two measures of goodness-of-fit are the *R-squared* (or adjusted R-squared) and the *Standard Error of Estimate*, SEE (Lewis-Beck and Skalaban1990). Take the regression model, say $Y = a + bX + cZ + dQ + e$, estimated (OLS) over a 12-election series. The higher the R-squared, the higher the linear fit of the model to the sample data. The lower the SEE, the greater the accuracy, on average, of predictions from the model. In general, these two measures move in tandem—the higher the R-squared, the lower the SEE. But while the relationship between the two is strong, it is not perfect. That says that R-squared maximisation, in itself, will not guarantee lower SEE. When one must choose one of these standards to pursue, it pays to concentrate on lowering the SEE, since it directly assesses estimation of Y.

Suppose the regression model forecasts Labour at 46 per cent in 2005, with SEE = 3.0. The SEE suggests the average error, when predicting upcoming elections. What is the likelihood that something close to this *point forecast* of 46 per cent will occur? A formal *confidence interval* (CI) is needed. The conventional two-tailed 95 per cent confidence interval would be [46 +/– 2.3 (3.0)]. In other words, the *interval forecast* for Labour is about [39, 53].

When confidence intervals are used to assess accuracy, the width of the band is sometimes discouraging, in that so many values are possible. In the case above, for example, the lower band value suggests a loss for the ruling party and the upper band value suggests a win. One approach to this problem is application of a *one-tailed* CI. The notion is that what is important is not exaggerating the margin of government victory, but rather gauging the likelihood of their defeat. In other words, the worry is not over whether actual Y is greater than forecast Y, but whether it is smaller, so risking defeat. In this case, a one-tailed CI would be [45 – 1.9 (3)], which gives 95 per cent certainty that the ruling party share is not less than 40. Note that this raises the lower bound, over the two-tailed test. We now have somewhat more confidence in the minimum size of the government's share.

Given the prevailing votes-to-seats ratio, a 40 per cent popular vote share could well translate into a government win.

The SEE, valuable as it is, has been criticised as too conservative an estimate of forecasting error (Beck 2000, 162–163). Strictly speaking, the SEE suggests the average prediction error, from a number of forecasts made when the X values are set at their means. When we wish to forecast an individual observation whose X values are not at the mean, as is typically the case with elections, the true forecast standard error will be greater than the SEE. The formula for the standard error of a particular forecast is somewhat cumbersome, and is specific to each case (Kmenta 1997, 248–252). Basically, it adjusts SEE by the distances of the X from means. Most election forecasters do not report the particular forecast standard error, relying instead on the SEE. This is justified, first, because it provides a good 'average' guess about the error; second, because it is easier to calculate; third, the calculation of true forecasting errors usually yields little gain in precision; and fourth, the confidence bands around the predicted Y, even using the less conservative SEE, are still uncomfortably large.

An approach to accuracy that is gaining ground explores after-the-fact 'out-of-sample' prediction properties of the regression model. Typically, each known election is omitted from the data set, the model re-estimated on the remaining observations and the omitted election 'forecast' from that re-estimated model. The absolute errors from these forecasts can be summed, and averaged. The *Mean Absolute Error* (MAE) gives a rough idea of how much real error the model generates, at least after-the-fact (see Campbell 2000, 27–29). Mathematically, the measure becomes more tractable if the forecast errors are squared, averaged, then the square root taken. This measure, the *Root Mean Squared Error* (RMSE), is also known as Theil's U (Lewis-Beck and Tien 1996; Wood and Park 2004). It suggests average forecasting error from multiple predictions, and generally yields conclusions not dissimilar from the simpler MAE.

The out-of-sample approach has variants. One might be called a 'largest-error' method, the other a 'step-ahead' method. The largest error method looks at all the above calculated out-of-sample errors, and uses the largest one as an estimate of maximum possible error. That maximum possible error becomes a benchmark for deciding how likely it is that the current forecast would incorrectly pick the winner. For example, James Campbell (2000, 29) found that, out of 12 elections, the largest out-of-sample error was 3.8 per cent. Hence, he concluded, if the current incumbent forecast was 53.9 per cent or more, there was a bit less than 1 chance in 12 that the incumbent would lose (with <50 per cent). One difficulty with this strategy is that we in fact know little about the distributional properties of these errors. Another difficulty is that the largest error may be an outlier, in which case the test could be unduly severe.

The second variant, the 'step-ahead' method, is familiar in more standard time series work. Say Y represents the future election outcome at time t + 1, and Y′ represents the forecast at time t. We may estimate the model on the entire time series data set, up to and including t, but not including t + 1. We then forecast a 'step-ahead', at election t + 1. Say we estimated a UK model from 1955–1997, then took one step ahead, to forecast 2001. We could continue backwards, re-estimating with an ever smaller series, estimate from 1955 to 1992 and forecast to 1997, and so on back in time (Lewis-Beck and Tien 2004). Such a strategy reveals something about model stability, and problematic outlier elections. Unfortunately, this data set is too small to allow more sophisticated time series work.

These four components of forecasting instrument evaluation—accuracy, lead, parsimony, reproducibility—can be brought together in an overall quality index, as follows (Lewis-Beck 1985a, 60–61):

$$Q = \frac{(3A + P + R)(L)}{M}$$

where Q = a quality index of the forecasting instrument, A = accuracy (rated 0–2), P = parsimony (rated 0–2), R = reproducibility (rated 0–2), L = lead time (rated 0–2), and M = the highest possible score for the numerator, 20 (to allow Q a theoretical upper limit of 1).

Note that accuracy (A) has three times the weight of parsimony (P) and reproducibility (R), in order to recognise its special importance. Note further that lead (L) is entered interactively. This allows it greater effect, including the ability to reduce Q to '0', should L = 0. A model can be scored '0' for lead when, for example, the model is not really estimable until after the election. This has sometimes occurred when strictly explanatory models are gerry-rigged for forecasting, say, by estimating what the X values, available after the election, might be. Such conditional exercises are to be avoided, when the goal is forecasting. A model that was highly accurate (2), very parsimonious (2), easily reproducible (2), with good lead time (2) would score Q = 1.0. A model with no lead (0), or with poor accuracy (0), little parsimony (0) and poor reproducibility (0) would score Q = 0.0. For most models, the Q index falls between 1.0 and 0.0.

Forecast Model Revision: A Theory-Driven Process

In the competition of rival forecasting models, some will show better quality than others. From one election to the next, model performance may vary. But over time, certain models will emerge as consistently superior. Why? I would argue the essential reason is theory (Lewis-Beck and Rice 1992, 141). 'Forecasting requires more than curve fitting. It wants good theory.' (Lewis-Beck and Tien 2000, 98) A winning forecasting model is based on sound evidence from voting behaviour studies. The forecaster, in formulating a model of election outcomes, should include variables that are well-measured proxies for individual voter decisions, as they aggregate. Certain explanatory variables will, of necessity, be excluded from the specification. This is not a problem, at least as long as they are not significantly correlated with the variables included. What counts is that the variables in the equation are based on strong electoral theory. This is particularly important given the small sample sizes we must work with.

Not everyone agrees with this theoretical perspective. Some believe that increased accuracy requires the effacing of theory. Campbell (2000, 182), for example, asserts that 'There is no reason to forecast with one hand tied behind your back in a mistaken belief that a good forecasting model must also be a good explanatory model'. In his view, inclusion of near-tautological variables—'too conceptually close to the dependent variable itself'—might be justified, if they improve accuracy (Campbell 2000, 182). One difficulty with this approach, aside from the specification issue, is that lead time also gets eroded, as the forecaster moves closer to placing the dependent variable on the right-hand side of the equation. Note that Campbell's (2004b, 734) model, for US presidential elections, generally carries less lead time than its rivals.

Not infrequently, a forecaster, after a few real election trials, seeks to revise the model specification. Indeed, reading the recent set of papers forecasting the US 2004 election, one is hard-put to find a model that was not changed at least somewhat in response to a past trial (see *PS: Political Science and Politics*, 37:4, 2004). For example, Alan Abramowitz (2004, 745), after asserting 'I have made no changes in the basic model', goes on to note: 'I have slightly modified the presidential approval measure'. Campbell, for one, argues against this strategy, saying that the same specification should be tested over many trials. 'Model stability (the constancy of model specification from one election to the next) must be a goal of election forecasting along with prediction accuracy and lead time before the election' (Campbell 2004a, 735).

The difficulty with the model stability perspective is that, first, it ignores what has been learned (about the specification) from errors in recent trials. Second, it amounts to a counsel of glacial model change since, over the course of an adult forecaster's lifetime (say 40 years), he or she could expect to receive test results for only 8–10 national elections. Third, it may be that the measure for the variable needs to change. That is, conceptually we may think the variable should be in the model, but we realise that our current measure of it is flawed. (This was the case with Lewis-Beck and Tien (2004) who dropped a flawed measure of prospective economic voting.) With respect to model revision, I favor the KISS strategy advocated by Arnold Zellner: Keep It Sophisticatedly Simple (García-Ferrer 1998). That is, judiciously introduce one or two theoretically potent new variables (or new measures of old variables), when the cumulating evidence suggests it.

How the US Models Have Fared

The 1984 presidential election offered the first opportunity for published comparison of forecasts from different statistical modellers (Lewis-Beck 1985a), but competitors were few. Three models reported a forecast range for Reagan from 51.3 per cent to 55.3 per cent, thus correctly picking him as winner but missing his landslide. Subsequently, the enterprise of election forecasting acquired momentum in political science. From 1992 on, about half a dozen teams of forecasters, usually but not always the same people, could be depended on to release pre-election presidential forecasts. Here is the forecast range for the incumbent two-party popular vote share in these contests: 1992 (Bush) 44.8 per cent–55.7 per cent; 1996 (Clinton) 54.8 per cent–58.1 per cent; 2000 (Gore) 53 per cent–62 per cent; 2004 (Bush) 49.9 per cent–57.6 per cent (Campbell and Garand 2000; Campbell 2004a).

First, we observe that forecasters tend to agree on the winner (i.e. the candidate who will gain >50 per cent). Further, they usually do pick the winner, with the notable exception of 2000. Here is the number of correct winner picks, by year: 1992, three of five; 1996, seven of seven; 2000, none of six; 2004, six of seven. But, picking the winner is a small part of the story. As Campbell (2004a, 733) sagely observes, 'their evaluation should be based on their success in predicting the vote [percentage]'. By this criterion, one sees that the models vary widely. At one extreme, 1992, the percentage point spread was 10.9. On average for this set of four elections, the range is 7.7 percentage points. This is a very wide spread. For example, a forecast of 49 per cent spells defeat, whereas a forecast of (49 + 7.7 = 56.7) means a landslide. (According to the vote–seat swing ratio, a 49 per cent two-party incumbent vote share converts to 40 per cent of the Electoral College vote; whereas a 56.7 per cent share converts to 78 per cent of the Electoral College vote; see the conversion formula in Lewis-Beck and Tien 2004, 757).

The average point estimate spread of 7.7 also happens to be the real point spread of forecasts for the 2004 US presidential election. That representative, and most recent, contest offers an excellent opportunity to evaluate the overall quality of the different models. In Table 1 are offered measures of our four evaluation components—accuracy, lead, parsimony and reproducibility. We begin with accuracy, the most important criterion. All (or almost all) of the models have reported an R-squared (or adjusted R-squared), a SEE and a MAE. Reporting of the other accuracy indices was slim to none. No one reported a root mean squared error, Theil's U or a two-tailed confidence interval. There was one report of a one-tailed confidence interval, of 'step-ahead' results, and a forecast error, respectively (Lockerbie 2004; Lewis-Beck and Tien 2004; Norpoth 2004). The 'largest error' out-of-sample was reported three times by, respectively, Campbell (2004b), Holbrook (2004) and Lockerbie (2004).

In column 1 of Table 1 the model is noted, along with the 2004 point forecast below. In columns 2–4 are the accuracy indicators—R-squared, SEE, MAE—for which we have rather complete data. In the 2004 presidential election President Bush received 51.2 per cent of the two-party popular vote. Does the model with the highest R-squared (that of LB at 0.94) generate the closest prediction? Yes, for an error of 1.3 (i.e. 51.2 − 49.9 = 1.3). The same story repeats itself with the SEE and, to a large extent, with the MAE. Model LB is best on these three criteria.

However, by the other criteria, Model LB yields more mixed results. First off, its forecast is on the wrong side of the 'win–lose' ledger, erroneously giving the win to Kerry. Further, in terms of lead, three models are clearly better (see column 5). Model N takes first, having generated its forecast an impressive nine months in advance. At the other extreme is Model C, making its forecast less than two months ahead. With respect to the criterion of parsimony, the models also exhibit considerable variation (see column 6). Here we simply measure parsimony by comparing the number of independent variables to the number of observations. By this standard, Model LB does the worst, with four independent variables on thirteen observations (4/13 = 0.31). Model C comes out on top, with only two independent variables on fourteen observations (2/14 = 0.14). In sum, Model LB appears half as parsimonious as Model C (0.14/0.31 = ½).

Table 1: Evaluation of US Presidential Election Forecast Models in 2004

Model	R-sq.	SEE	MAE	Lead	Parsimony	Reproduce	Quality
A (53.7)	0.90	2.0	<2.0	94	3/14	2	0.75
C (53.8)	0.91a	1.8	1.6	57	2/14	2	0.50
H (54.5)	0.88a	1.9	2.0	64	3/12	2	0.56
L (57.6)	0.87	2.5	2.5	165	2/12	2	0.70
LB (49.9)	0.94	1.5	1.3b	67	4/13	2	0.68
N (54.7)	0.92a	2.5	—	278	5/23	1	0.65
W (52.9)	0.71	3.2	2.2	67	2/13	1	0.45

Note: In column 1, each letter stands for the model of the modeller(s): A = Abramowitz (2004); C = Campbell (2004b); H = Holbrook (2004); L = Lockerbie (2004); LB = Lewis-Beck and Tien (2004); N = Norpoth (2004); W = Wlezien and Erikson (2004); the figure in parentheses is the model forecast of the incumbent two-party popular vote share (in a case where there was more than one forecast, here Wlezien and Erikson (2004), the one closest to the group mean of the single forecasts is reported. In column 2, an 'a' indicates an adjusted R-squared. In column 3, SEE = the standard error of estimate. In column 4, MAE = the mean absolute out-of-sample error; a '—' indicates none was reported; a 'b' indicates that it was calculated by averaging the reported step-ahead forecasts. In column 5, Lead = the number of days before the 2 November election the forecast was issued. In column 6, Parsimony = the number of independent variables divided by the sample size. In column 7, Reproduce = the score (0–2) on ability of another researcher to reproduce the model, as judged by my reading of measurement construction and data sources in the article where the model appeared. In column 8 Quality = the Quality Index, calculated from the formula provided in this article; here are the following numbers I entered, respectively, for Accuracy, Parsimony, Reproducibility and Lead—A = 1.5, 1, 2, 2; C = 2, 2, 2, 1; H = 1.5, 1, 2, 1.5; L = 1, 2, 2, 2; LB = 2, 1, 2, 1.5; N = 1.5, 1, 1, 2; W = 1, 2, 1, 1.5.

The last criterion, reproducibility, is perhaps the most subjective (see column 7). I carefully reviewed the construction of each model, and then judged how easy it would be to achieve the same results, on the basis of what I had read. My conclusion is that I would have some difficulty with Model N and Model W. I did not feel it would be impossible, but I doubted my ability to do it handily, and to achieve the same results. Therefore, I graded them down a bit ('1' instead of '2'). Of course, readers may judge for themselves and see if they agree.

How do the results from these criteria, combined together in the Quality Index developed earlier, rank these different models overall? These Q scores are reported in column 8. (More details on their construction appear at the bottom of the table.) The findings are non-obvious. The top is Model A, with Q = 0.75. At the bottom is Model W, with Q = 0.45. Model W's lower score is driven by its relatively modest marks on the R-squared, the SEE and the MAE. Also, it is middling with regard to lead and reproducibility. By way of illustration, it is worth presenting Model A:

Model A, by Abramowitz (2004), is as follows in this OLS regression:

$$V = 50.75 + 0.107^{*}P + 0.818^{*}E - 5.14^{*}T + e \qquad \text{(Eq.4)}$$
$$(0.025) \quad (0.183) \quad (1.24)$$
$$\text{R-squared} = 0.90 \quad \text{SEE} = 2.00 \quad N = 14 \text{ (elections 1948–2000)}$$

where V = the incumbent's share of the two-party popular vote, P = presidential approval minus presidential disapproval in the last June Gallup poll, E = annualised growth rate of real GDP in the first two quarters of the election year, T = a term dummy, scored '1' if the president's party has been in the White House for more than one term, '0' otherwise; the figures in parentheses are standard errors, * = statistically significant.

Abramowitz (2004) has used this 'time-for-a-change' model, with slight modifications, to forecast every election since 1988. He has correctly predicted the popular vote winner in each of these elections. His 2000 forecast of 53 per cent for Gore was closer than most other forecasts for that contest (Campbell and Garand 2000). His 2004 forecast, 53.7 per cent for Bush, was off by just 2.5 percentage points. Note how closely his model follows what I call the core specification (see Eq. 2), adding to it the institutional, term variable. In other words, the model is built on 'strong theory'. Also note the relative simplicity of the variables, conceptually and in terms of their measurement. These things, coupled with its enviable track record over the years, suggest it is emerging as a leader in the forecasting business.

US Models Compared to Other Methods

Thus far we have evaluated statistical models themselves, compared to other statistical models. It is now appropriate to compare them to alternative approaches. In 2004, the forecast error for Model LB, which was closest, = 1.3 (i.e. 51.2 – 49.9). Taking the models of Table 1 as a group, their average forecast = 53.9, for an error of 2.7. (i.e. 53.9 – 51.2 = 2.7). The major rival approach comes from the vote intention surveys, with Gallup the chief contender. Looking at the Final Gallup Pre-election Surveys, the average prediction error for the incumbent two-party popular vote share over the period 1948–2000 = 2.1 percentage points. However, the Final Gallup Pre-Election forecast for 2004 was 49 per cent—Kerry, 49 per cent—Bush, 1 per cent—Nader (Gallup Poll, 1 November 2004). That is to say, they saw Kerry and Bush splitting the two-party popular vote 50–50. This gives Gallup a prediction error of 1.2 for 2004 (i.e. 51.5 – 50 = 1.5).

The other major approaches, being newer, have less of a before-the-fact track record. Still, we can report, on the basis of a Pew Research Center Final Pre-election Survey (27–30 October 2004), that the 'citizen forecast' gave an expected vote to President Bush of 52.9, for an error of 1.7 (Lewis-Beck and Tien (forthcoming)). Finally, the Iowa Electronic Market (IEM) forecast 50.5 for Bush, on 1 November, for an error of 0.7 percentage points.

One observes that, in 2004, each of the scientific approaches to forecasting did rather well, with an error of 1–2 points or so. All approaches manifest considerable improvement over a baseline, know-nothing guess, where the prediction is like a coin toss, giving the incumbent candidate a 50 percentage point share. Over the 1948–2004 period, this know-nothing guess generates a prediction error of 4.6 percentage points. For any given year, of course, the different scientific approaches may not fare the same, and may even do worse than the know-nothing guess. The 2000 US presidential election stands out here. For that contest, the leading statistical models all saw an easy Gore victory (53–60 per cent), when in fact he lost with 50.3 per cent of the two-party popular vote (Lewis-Beck 2001). Gallup's last estimate, published on election day 2000, saw Bush ahead in the popular race by about two points (Bush = 48, Gore = 46 and Nader = 4). The IEM's election-day forecast saw Bush ahead, plus 3 points. Forecasters from other academics, regardless of method employed, virtually all forecast a Gore victory. For example, in an international survey of expert opinion, 49 out of 50 academics forecasted Gore (*Le Figaro*, 20 October 2000, xiii).

Much was learned from the US presidential election of 2000 with respect to the world of election forecasting. It demonstrated that, in a particular contest, all types of methods can yield poor results; furthermore, most examples from a single type of method can yield an especially poor result. In 2000, forecasters did a bad job, especially statistical modellers. But, in 1992, the polls did quite badly. For example, in 1992, the Gallup final pre-election survey missed by 5.8 percentage points. However, in 1996, and then in 2004, statistical modellers overall did a good job (see Campbell and Garand 2000).

The message is that it is dangerous to reject wholesale a forecasting method on the basis of one election trial. When the track record of the two major competitors—models and polls—are compared over the long haul, it seems clear that on average they deliver about the same accuracy, with something like an average error of a little over two points. In addition, and importantly, the models have the advantage of lead, in that they speak several months before the contest, while the polls do not declare until a day or two before. Statistical models, then, are capable of substantive, as opposed to merely symbolic, forecasts, and as such are to be preferred.

How French Models have Fared

Outside of the US, forecasting from statistical models is most extensive in France. The French case has special value, in that it shows the possibilities of forecasting in an ostensibly rather different democratic system, a presidential-parliamentary hybrid with multiple parties, complex coalition strategies and two rounds of balloting. The work comes out of popularity functions (PF) and vote functions (VF). The PF have as the dependent variable 'satisfaction' with the president (or prime minister) as measured in a national survey, while the VF have as the dependent variable presidential (or legislative) vote share. (See the excellent review essay by Bélanger (2004).) The first forecasting paper was by Jean-Dominique Lafay (1977), who estimated a popularity function to forecast the 1978 legislative contest. The second forecasting paper was by Lewis-Beck (1985b), who estimated a vote function to forecast the 1986 legislative contest. Subsequent efforts have been mostly on vote functions, appropriately modified to permit before-the-fact forecasts (Lewis-Beck 1995 and 2002).

The dependent variable for these vote functions may be the incumbent (coalition), or the left coalition. The data are national time series, or pooled time series cross-sections from the regions or the departments. The chief, not to say only, rival to the statistical modelling approach here is the polls. In 1997, the several commercial polling firms foresaw an easy victory of the ruling right coalition in the National Assembly contest. However, these polls were wrong, and the statistical models were right, forecasting the coming to power of the left (Fauvelle-Aymar and Lewis-Beck 1997; Jérôme et al. 1999).

In the 2002 elections, forecasters of all stripes, scientific or not, failed to foresee the first-round defeat of Jospin, and the rise of Le Pen. With respect to the pre-election polls just before that April contest, average error was about six percentage points. For IFOP, perhaps the most widely known firm, this average error rose to 7.7 per cent (Jérôme and Jérôme-Speziari 2004, 182–185). The models also did badly. The French-named 'Iowa Model' incorrectly forecast a narrow second-round victory for Jospin (Fauvelle-Aymar and Lewis-Beck 2002). After-the-fact forecasts for 2002 were rendered by Bruno Jérôme and Veronique Jérôme-Speziari (2004) and by Eric Dubois and Christine Fauvelle-Aymar (2004), in pooled analyses. For example, in the former, the after-the-fact forecast was 42.4 per cent of the first-round presidential vote for the Plural Left, which actually received 42.9 per cent (Jérôme and Jérôme-Speziari 2004, 192).

Dubois and Fauvelle-Aymar (2004, 218) made a step-ahead prediction of the first-round 2002 legislative contest (from a pooled regional data set 1986–1997). It gave the Left 43.3, when in fact it got 41.2, for an error of 2.1 percentage points. For the model, the SEE = 1.13, which suggests that this 2002 projection was within normal bounds. It is perhaps useful to present a version of their model, by way of illustrating the French work (Dubois and Fauvelle-Aymar 2004, 215–216). The basic model, below, follows the general core political economy model (see Eqs. 1–3), as do most of the French forecasting equations:

$$V = a + bP + cE + e \qquad \text{(Eq.5)}$$

where V = vote share received by left parties at the first round of the legislative election, P = popularity of the left parties according to SOFRES polls three months before the election, E = regional unemployment rate in the quarter preceding the election quarter. Here are the model estimates (random effects):

$$V = 35.97^{*} + 8.44^{*}P - 0.17^{*}E + e \qquad \text{(Eq.6)}$$

$$\text{Adj. R-squared} = 0.81 \quad N = 110 \text{ (regions from 1986–2002)}$$

The accuracy of this (and other French models) as judged by point forecasts, the R-squared and the SEE are not dissimilar to the US models. (None of the French work looks at other measures of accuracy.) Furthermore, this model fares well in terms of the remaining forecast evaluation criteria of lead, parsimony and reproducibility.

One difficulty it shares with other French efforts is conversion of vote forecasts to seat forecasts (Dubois and Fauvelle-Aymar 2004, 219; Jérôme and Jérôme-Speziari 2004, 194–195). The swing-ratio equation predicting final seat total from first-round vote share (regional level) yields a moderate fit (adj. R-sq. = 0.60) and an average within-sample error of 27 seats. In some cases, the prediction is way off, as in 1993 when the predicted seats were 147, but the actual were 82. Dubois and Fauvelle-Aymar (2004, 223) conclude that perhaps it would be better to have the dependent variable as seats, rather than votes, thereby eliminating the necessity of a swing-ratio. It is noteworthy that this swing-ratio problem is less great in the US case, for presidential two-party popular vote is highly predictive of Electoral College vote, and congressional forecasting models usually use seats themselves as the dependent variable (Lewis-Beck and Bardwell 2004; Lewis-Beck and Tien 2004). However, for the British case, to which we now turn, the swing-ratio issue is again problematic.

The British Case: Context and Conclusion

While statistical models of election forecasting have been pursued in the UK, the enterprise has been relatively less developed than in the US and France. The US work, and plenty of it, has concentrated on time series vote functions. For France, the body of work is growing, having moved quickly from popularity functions to vote functions, with time series (or pooled time series) data. For the UK, the focus has been on popularity functions. In particular, the effort has been to forecast the vote intention for government, as measured in public opinion polls over time. Paul Whiteley (1979) made the first attempt, with a Box-Jenkins modelling of this monthly time series on popularity. David Sanders (1991) continued the popularity function focus, introducing substantive independent variables as well.

However, in all this popularity work, what was forecast was vote intention. The question might be: ‘Can a model forecast the next vote intention score in the absence, or presumed absence, of a vote intention poll at that future time point?’. This sets up a difficult test, in which the aim might be to beat the polls at their own game. But, all the time, what is forecast is vote *intention*, not vote *share*. This necessarily introduces error, barring the unlikely event that vote intention equals vote share. Further, it is not forecasting what really counts—votes. The first scholar to recognise this difficulty in print was Anthony Mughan (1987), who estimated some simple vote functions.

Unfortunately, his call for further effort on vote functions was not taken up until Lewis-Beck et al. (2004). That paper formed the basis of a special British election forecasting symposium in *Electoral Studies* (23, 279–313). Lewis-Beck et al. (2004) offered the following regression equation (OLS):

$$V = {}^{*}{-42.7} + {}^{*}0.27P - {}^{*}0.97E - {}^{*}3.1T + e$$

$$\text{R-sq.} = 0.88 \quad \text{adj. R-sq.} = 0.83 \quad \text{SEE} = 2.27 \quad N = 12 \text{ (observations 1955–1997)}$$

where V = government party vote share, P = public approval of the government record six months prior, E = inflation rate six months prior.

Model specification follows the core political economy specification employed in US and French models (see Eqs. 1–3). In particular, it appears much like the Abramowitz formulation for US presidential elections (see Eq. 4 above). With respect to accuracy measures, the R-squared are high, the SEE low. The mean absolute out-of-sample forecasting error (MAE) = 2.3, almost exactly the same as the SEE. Other diagnostics suggest that model has considerable stability. That model was used to issue a before-the-fact forecast of the 2001 Labour vote share, which was then translated into a seat share. Discussion of that forecast, and subsequent model tests, form the basis of one of the articles in this issue of the journal.

Given all this related work, we now have an emerging UK competition. Statistical modellers will compete among themselves, and with the pollsters (and possibly other approaches), for the better forecasting instruments. Questions to be answered are the same as elsewhere. Which forecasting instrument, from which approach, has more accuracy, lead, parsimony and reproducibility? In the UK, as in the US and France, the expectation is that the statistical modellers will be able to do at least as well as the competition, if not better.

About the Author

Professor Michael S. Lewis-Beck, Department of Political Science, University of Iowa, Iowa City, Iowa 52242, USA, email: *michael-lewis-beck@uiowa.edu*

Note

1. A frequent objection to the specification of US presidential forecasting models is that they include no campaign or candidate variables. Work has shown that these specific additions do not add significantly to the models (Lewis-Beck and Tien 2002; Nadeau and Lewis-Beck 2001). Besides that, in some senses, the campaign is actually taken into account in these models. First, they may assume that the campaign is in equilibrium. That is, each major-party candidate campaigns sincerely and for the same amount of time, thus effectively cancelling out the effects of the other. Second, one may consider that campaign/candidate variables are in the error term, and their absence in the equation explains the average two per cent or so error of these models. Third, these models, as will be seen, are about as accurate as final pre-election day polls, which have had the benefit of the entire campaign season. Again, that speaks to the notion that, effectively, campaign forces are taken into account by the models.

Bibliography

Abramowitz, A. (2004) 'When good forecasts go bad: the time-for-change model and the 2004 presidential election', *PS: Political Science and Politics*, 37:4, 745–746.

Armstrong, S. (ed) (2001) *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Boston: Kluwer Academic Publishing).

Bean, L. (1948) *How to Predict Elections* (New York: Knopf).

Beck, N. (2000) 'Evaluating forecasts and forecast models of the 1996 presidential election', in J. Campbell and J. Garand (eds), *Before the Vote: Forecasting American National Elections* (Thousand Oaks: Sage).

Bélanger, E. (2004) 'Finding and using empirical data for vote and popularity functions in France', *French Politics*, 2, 235–244.

Brody, R. and Sigelman, L. (1983) 'Presidential popularity and presidential elections: an update and extension', *Public Opinion Quarterly*, 47, 325–328.

Campbell, J. (2000) 'Polls and votes: the trial-heat presidential election forecasting model, certainty, and political campaigns', in J. Campbell and J. Garand (eds), *Before the Vote: Forecasting American National Elections* (Thousand Oaks: Sage).

Campbell, J. (2004a) 'Introduction—the 2004 presidential election forecasts', *PS: Politics and Political Science*, 37:4, 733–736.

Campbell, J. (2004b) 'Forecasting the presidential vote in 2004: placing preference polls in context', *PS: Politics and Political Science*, 37:4, 763–768.

Campbell, J. and Garand, J. (eds) (2000) *Before the Vote: Forecasting American National Elections* (Thousand Oaks: Sage).

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

Fair, R. (1978) 'The effect of economic events on votes for president', *Review of Economics and Statistics*, 64, 159–172.

Fauvelle-Aymar, C. and Lewis-Beck, M. (1997) 'L'Iowa donne l'opposition gagnante', *Libération* (4978, 23 May), 15.

Fauvelle-Aymar, C. and Lewis-Beck, M. (2002) 'Pour l'Iowa, avantage Jospin', *Libération* (6485, 21 March), 6.

Forsythe, R., Nelson, F., Neumann, G. and Wright, J. (1989) 'The Iowa presidential stock market: a field experiment', in R. Issac (ed), *Research in Experimental Economics, Volume 4* (Westport: JAI Press).

García-Ferrer, A. (1998) 'Professor Zellner: an interview for the International Journal of Forecasting,' *International Journal of Forecasting*, 14:3, 303–312.

Hibbs, D. (1982) 'President Pragan's mandate from the 1980 elections: a shift to the right?', *American Politics Quarterly*, 10, 387–420.

Holbrook, T. (2004) 'Good news for Bush? Economic news, personal finances, and the 2004 presidential election', *PS: Political Science and Politics*, 37:4, 759–762.

Holly, S. and Weale, M. (eds) (2000) *Econometric Modelling: Techniques and Applications* (Cambridge: Cambridge University Press).

Jérôme, B., Jérôme, V. and Lewis-Beck, M. (1999) 'Polls fail in France: forecasts of the 1997 legislative election', *International Journal of Forecasting*, 15:2, 163–174.

Jérôme, B. and Jérôme-Speziari (2004) 'Forecasting the 2002 elections: lessons from a political economy model', in M. Lewis-Beck (ed), *The French Voter: Before and After the 2002 Elections* (London: Palgrave Macmillan).

Kmenta, J. (1997) *Elements of Econometrics* (2nd edn) (Ann Arbor: University of Michigan Press).

Lafay, J-D. (1977) 'Les consequences électorales de la conjuncture économique: essai de prevision chiffrée pour mars 1978', *Vie et Sciences Economiques*, 75, 1–7.

Lewis-Beck, M. (1985a) 'Election forecasts in 1984: how accurate were they?', *PS: Political Science and Politics*, 18:1, 53–62.

Lewis-Beck, M. (1985b) 'Un modèle de prevision des élections legislative Françaises (avec une application pour 1986)', *Revue Française de Science Politique*, 35:6, 1080–1091.

Lewis-Beck, M. (1995) 'Comparaison de prevision des elections présidentielles en France et aux Etats-Unis', *Journal de la Société de Statistique de Paris*, 136:1, 29–45.

Lewis-Beck, M. (2001) 'Modelers v. pollsters: the election forecasts debate', *The Harvard International Journal of Press and Politics*, 6:2, 10–14.

Lewis-Beck, M. and Bardwell, K. (2004) 'State-level forecasts of US Senate Elections', *PS: Political Science and Politics*, 37:4, 821–826.

Lewis-Beck, M., Nadeau, R. and Bélanger, E. (2004) 'General election forecasts in the United Kingdom: a political economy model', *Electoral Studies*, 23, 279–290.

Lewis-Beck, M. and Rice, T. (1982) 'Presidential popularity and presidential vote', *Public Opinion Quarterly*, 46: Winter, 534–537.

Lewis-Beck, M. and Rice, T. (1984) 'Forecasting presidential elections: a comparison of naïve models', *Political Behavior*, 6:1, 9–21.

Lewis-Beck, M. and Rice, T. (1992) *Forecasting Elections* (Washington, DC: Congressional Quarterly Press).

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

Lewis-Beck, M. and Skalaban, A. (1990) 'The R-squared: some straight talk', *Political Analysis*, 2, 153–171.

Lewis-Beck, M. and Tien, C. (1996) 'The future in forecasting: prospective presidential models', *American Politics Quarterly*, 24:4, 468–491.

Lewis-Beck, M. and Tien, C. (1999) 'Voters as forecasters: micromodels of presidential election prediction', *International Journal of Forecasting*, 15:2, 175–184.

Lewis-Beck, M. and Tien, C. (2000) 'The future in forecasting: prospective presidential models', in J. Campbell and J. Garand (eds), *Before the Vote: Forecasting American National Elections* (Thousand Oaks: Sage), 133–144.

Lewis-Beck, M. and Tien, C. (2002) 'Presidential election forecasting: the Bush-Gore draw', *Research in Political Sociology*, 10, 173–187.

Lewis-Beck, M. and Tien, C. (2004) 'Jobs and the job of president: forecast for 2004', *PS: Politics and Political Science*, 37:4, 753–758.

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

Lockerbie, B. (2004) 'A look to the future: forecasting the 2004 presidential election', *PS: Political Science and Politics*, 37:4, 741–744.

Mughan, A. (1987) 'General election forecasting in Britain: a comparison of three simple models', *Electoral Studies*, 6, 195–207.

Nadeau, R. and Lewis-Beck, M. (2001) 'National economic voting in US presidential elections', *Journal of Politics*, 63:1, 159–181.

Norpoth, H. (2004) 'From primary to general election: a forecast of the presidential vote', *PS: Politics and Political Science*, 37:4, 737–740.

Rosenstone, S. (1983) *Forecasting Presidential Elections* (New Haven: Yale University Press).

Sanders, D. (1991) 'Government popularity and the next general election', *Political Quarterly*, 62, 235–261.

Sigelman, L. (1979) 'Presidential popularity and presidential elections', *Public Opinion Quarterly*, 43, 532–534.

Tufte, E. (1978) *Political Control of the Economy* (Princeton: Princeton University Press).

Whiteley, P. (1979) 'Electoral forecasting from poll data: the British case', *British Journal of Political Science*, 9, 219–236.

Wlezien, C. and Erikson, R. (2004) 'The fundamentals, the polls, and the presidential vote', *PS: Political Science and Politics*, 37:4, 747–752.

Wood, B. and Park, S. (2004) 'Prediction equation', in M. Lewis-Beck, A. Bryman and T. Liao (eds), *The Sage Encyclopedia of Social Science Research Methods* (Thousand Oaks: Sage), 851–852.