Opinion polling and election predictions

Abstract
Election prediction by means of opinion polling is a rare empirical success story for social science. I examine the details of a prominent case, drawing two lessons of more general interest:
1) *Methodology over metaphysics*. Traditional metaphysical criteria were not a useful guide to whether successful prediction would be possible; instead, the crucial thing was selecting an effective methodology.
2) *Which methodology?* Success required sophisticated use of case-specific evidence from opinion polling. The pursuit of explanations via general theory or causal mechanisms, by contrast, turned out to be precisely the wrong path – contrary to much recent philosophy of social science.

1. Introduction: metaphysics and methodology

Is systematic predictive success in social science possible? Many have given reasons why it is not, such as the fact that social systems are open systems, or that they exhibit reflexivity, or simply that there are too many variables needing to be modeled (Taylor 1971, Giddens 1976, Hacking 1995, Lawson 1997). In this paper I examine a notable case of predictive success so far relatively neglected by philosophers – namely election prediction by means of opinion polling – that seems to contradict these reasons.

Next, if successful prediction is possible, what makes that so? The lesson from the opinion polling case is that the most fruitful answer to this question is not metaphysical but rather is *methodological*. In particular, success or the lack of it was not predictable from the metaphysics of elections, which indeed in many respects remain unknown.[1] Rather, what was crucial was a certain methodological approach.

One popular methodological view, borne in part from pessimism about the possibility of prediction, has argued that the main aim of social science should instead be *explanation*.

---

[1] Strevens (2005), for instance, gives metaphysical conditions for when explanation and prediction are possible in some complex systems. But it is unclear whether these conditions are satisfied in the elections case.

This latter can be achieved via the discovery of causal mechanisms, as urged by 'new mechanists' (Lawson 1997, Brante 2001), or else via the development of underlying theory (Elster 1989, Little 1991). Moreover, much of mainstream practice in economics and other social sciences is implicitly committed to this latter view: while all accept that rational choice models, for instance, might not be predictively successful, nevertheless they are held to provide 'understanding' or 'underlying explanation'.

A contrary view rejects this methodological emphasis on mechanisms or underlying theory (Cartwright 2007, Reiss 2008). One strand, motivated in part by detailed case studies of other empirical successes, has emphasized instead context-specific and extra-theoretical work. Theories and mechanisms play at most a heuristic role; empirical success requires going beyond them (Alexandrova 2008, Alexandrova and Northcott 2009).

The details of the opinion polling case turn out to endorse this second view. The reason is that, roughly speaking, while the extra-theoretical approach achieved prediction but not explanation, the theory-centred approach achieved neither. That is, a search for explanation not only yielded no predictive success, it also yielded no explanations. So the first view gives exactly the wrong advice.

I focus on the 2012 US presidential election, in which Barack Obama defeated Mitt Romney. I begin by describing the predictive success achieved by aggregators of opinion polls (section 2), before examining how this success was achieved (sections 3 and 4). In contrast, theoretical approaches to election prediction fared much worse (section 5). I then discuss their failure also at furnishing explanations (sections 6 and 7).

2. Predictive success and metaphysical criteria

The 2012 presidential campaign featured literally thousands of opinion polls. The most successful of all election predictors were some *aggregators* of these poll results. Famously, several correctly forecast the winner of all 51 states in the 2012 election, as well as also getting Obama's national vote share correct to within a few tenths of a

percent.[2] This was a stunning success, arguably with few equals in social science. Nor was it easy – no one else replicated it, although many tried.[3] On the morning of the election the bookmakers had Romney's odds at 9/2, i.e. about 18%. Political futures markets such as InTrade had Romney's chances at about 28%. These market prices imply that common opinion was surprised by the outcome.[4]

Moreover, it is unreasonable to declare this success a mere fluke. First, the same poll aggregators have been successful in other elections too. And within any one election there have been many separate successful predictions, such as of individual Senate races or of margins of victory, which are at least partially independent of each other. Second, the aggregators' methods are independently persuasive. It therefore behooves philosophers of social science to understand them.

Meanwhile, do elections satisfy the metaphysical criteria allegedly necessary for predictive success? It seems not. Presidential elections are clearly open systems in that

---

[2] Four of the most successful aggregators were: http://votamatic.org, http://www.huffingtonpost.com/news/pollster/, http://election.princeton.edu/, and http://fivethirtyeight.blogs.nytimes.com/. The forecasting models for the first two of these were designed mainly by political science academics, the third by a neuroscience academic, and the last by a non-academic. Three of the four got every state right.

[3] See http://delong.typepad.com/sdj/2012/11/war-on-nate-silver-final-after-action-report-the-flag-of-science-flies-uncontested-over-silvergrad-weblogging.html for a list of 47 examples of failure, with an emphasis on their suspicion of polling-based prediction.

[4] There is an issue here about exactly what we are predicting and, thus, how we measure predictive success. After all, these market prices still had Obama as favorite, so why should we term them 'surprised' by his victory? In reply, first, besides the simple fact of the overall winner, there were also relevant additional facts: who won each state; and how much they won them by. Odds-makers were not impressive with respect to these more detailed targets. Indeed, barring unlikely background assumptions, the details of the state-level results are hard to reconcile with a 28% chance of overall Romney victory. Second, the best poll aggregators' predictions were probabilistic, which makes it quite an intricate matter assessing who actually did best. (For analysis, see http://rationality.org/2012/11/09/was-nate-silver-the-most-accurate-2012-election-pundit/.) But there is no serious dispute that the odds-makers and many other predictors were not accurate.

they are not shut off from causal influences unmodelled by political science. They undoubtedly feature many variables. And they are clearly subject to reflexivity concerns: sometimes the mere publication of a poll itself influences an eventual election result; indeed there were several examples of this during the 2012 campaign. Yet despite such troubles, highly successful prediction proved possible nevertheless.

3. The science of opinion polling

In any opinion poll, the voting intentions of a sample serve as a proxy for those of a population.[5] How might things go wrong, such that the sample will not be representative? The most well-known way, a staple of newscasts and introductory statistics courses alike, is *sampling error*: small samples can lead to misleading flukes. But sampling error is not the only, or even the most important, source of inaccurate predictions.[6] Awareness of this crucial point lies at the heart of any serious election prediction.

To begin, a major issue for pollsters is to ensure that their samples are appropriately balanced with respect to various demographic factors. Suppose, for example, that two-thirds of interviewees were women. Since there was good reason to think that women were disproportionately likely to vote for Obama, it follows that such a woman-heavy sample would give misleadingly pro-Obama predictions. Polling companies would therefore *rebalance* such a sample, in effect putting greater weight on men's responses. Notice several things about such a rebalancing procedure.

First, it is quite different from sampling error.[7] In particular, if our sampling procedure over-selects women, then this problem will not be alleviated just by making the sample larger.

---

[5] Although election prediction is the focus of this paper, opinion polls of course have many other uses too.

[6] For example, almost 25% even of late polls miss the final election result by more than their official confidence interval (http://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/), yet the expected miss rate given sampling error alone should be only 5%.

[7] Lying in the background here are reference class issues. But given the unavoidable cognitive and epistemic constraints facing polling scientists, their choice of reference class is not arbitrary. And in practice the distinction between sampling and systematic error is of enormous importance.

Second, sample rebalancing is clearly unavoidable if we wish to predict accurately. For this reason, every polling company performs some version of it.

Third, a poll's headline figures are therefore heavily *constructed*. They are certainly not the raw survey results. Exactly what and how much rebalancing is required depends on assumptions about the actual turnout on election day. For instance, in recent American presidential elections typically there have been slightly more women than men voters, so it would be a mistake to rebalance the sample to exactly 50-50. The correct figure may not be obvious, it needing to be inferred from imperfect polling data about past elections, and moreover with some assessment of how patterns of turnout may change again in the upcoming election. Accordingly, different polling companies may reasonably choose slightly different rebalancing procedures. The result is the phenomenon of 'house effects', when any particular company's results may systematically favor one or other candidate compared to the industry average. When assessing the significance of a poll for election prediction, it is vital to be aware of this.

Fourth, the rebalancing issue is pressing because it applies to many other factors besides gender, such as: age; income; race; likeliness to vote; education; ownership of cellphones but not landlines; and home access to internet. Not only is the precise rebalancing procedure for each of these factors arguable, it is also arguable exactly which factors should be rebalanced for in the first place (see below).

In addition to random sampling error and systematic sampling bias, there are several other potential sources of error as well. There is space only to mention a couple here. One is the phenomenon of herding: at the end of a campaign pollsters – it is widely suspected – 'herd', i.e. report headline figures closer to the industry mean, presumably to avoid the risk of standing out as having missed the final result by an unusually large margin. Some sensitivity to this turns out to be optimal for accurate election prediction.[8] A second

---

[8] For evidence of herding's significance, and references, see: http://fivethirtyeight.com/features/heres-proof-some-pollsters-are-putting-a-thumb-on-the-scale/,

worry is simply that voters may change their minds between a poll and election day. This is the main reason why polls taken, say, six months before an election have a much poorer predictive record than do those taken later. Election predictions must therefore take into account a poll's date too.

4. Poll aggregation

Turn now to the aggregation of polls, which represents a second layer of method, quite distinct from that required to conduct a single poll. Historically, poll aggregation has had a better predictive record than using individual polls alone. One obvious reason is that aggregation increases effective sample size and therefore reduces sampling error. A typical individual poll may have 95% confidence intervals of 3 or 4%; those for an aggregation of eight or ten polls, by contrast, are typically 0.75 or 1%.[9] But it is a different story for the other possible sources of error. Mere aggregation is no cure for those, because they may bias all polls – and hence the aggregate of polls – in a similar way.[10]

What, then, does account for aggregation's superior predictive success? In part, it is indeed simply the reduction of sampling error. But it is not just that; it is also that *sophisticated* aggregation can mitigate the other sources of error too. This explains why the best aggregators beat simple averaging of the polls. It is instructive to consider a couple of examples of methodological issues in more detail.

4.1) State versus national polls

One feature of the 2012 US presidential campaign was a divergence between state and national polls. By combining opinion polls for individual states, making due allowance

---

http://fivethirtyeight.blogs.nytimes.com/2012/10/11/oct-10-is-romney-leading-right-now/, http://votamatic.org/pollsters-may-be-herding/, and http://themonkeycage.org/2012/09/26/robo-polls-a-thumb-on-the-scales/.

[9] There were rarely more than eight or ten polls of a single area in a single time period.

[10] Although house effects will, by definition, cancel out, still it might be that the best sample rebalancing procedure is an outlier relative to the industry average.

for population and likely turnout, it is possible to calculate an implicit figure for the vote share across the country as a whole. When this was done, there was a surprising inconsistency: the state polls implied that Obama was ahead at the national level, but the national-level polls showed him behind. The divergence was at least three percentage points. What to do? Simple averaging was no answer, because the inconsistency was equally true of the polls' averages too.

One possible cause of the divergence was that it was just sampling error – confidence intervals are sufficiently wide that there is a non-negligible chance of this. However, the discrepancy had persisted for much of the campaign, rendering this explanation implausible. Another possibility was that Obama's votes were disproportionately concentrated in heavily-polled swing states. But this explanation turned out to be implausible too: first, it required disproportionately good Romney polling in non-swing states, but this had not occurred. Second, it seemed unlikely anyway given that, demographically speaking, swing voters in Ohio or Virginia are much the same as those in Texas or California, so why should their voting intentions be systematically different? – after all, both campaigns were spending similar amounts in the swing states. Third, such a pattern is uncommon historically.

There therefore seemed little prospect of reconciliation; instead, it boiled down to preferring one of the state or national polls to the other. In favor of the national polls, they tend to have larger sample sizes and to be run by more reputable firms. But on balance, there were better reasons to prefer the state polls. First, there are many more of them, suggesting that sampling error is less likely. Second, some of the other sources of error are arguably less likely too. In particular, herding effects will likely occur relatively independently in different states. As a result, that source of error for state polls will likely cancel out when inferring aggregate national polling numbers. Third, historical evidence again: when the two have conflicted in previous elections, typically the state polls have proven better predictors than have national ones.

The take-away is to emphasize the value added by sophisticated poll aggregation. Simply averaging the polls was not enough. Neither was it optimal just to split the difference between state and national polls symmetrically. Instead, more sophisticated analysis was required.

4.2) Could all the polls have been wrong?

By the end of the 2012 campaign, it was clear that if the polls were right then Obama would win. Romney's only hope by then was that the polls were systematically skewed against him. Thus, all turned on whether the polls could indeed be so skewed. Once again, simple averaging is no help here because the issue at hand is not *what* the polls said but rather how strongly we should *believe* what they said.

The historical record suggested it was unlikely the polls were skewed enough to save Romney.[11] Confidence in this conservative verdict was strengthened by the absence in 2012 of factors that had marked systematic polling errors in the past, such as a high number of voters declaring themselves 'undecided', or a significant third-party candidate. Given that it was the end of the campaign, there was also little reason to expect a large change of voters' minds before election day – especially given also the record levels of early voting. Finally, the number of polls involved and the size of Obama's lead made sampling error too an implausible savior for Romney.

The only remaining source of error was therefore sample rebalancing. In particular, was there some procedural skew, common across many or all polls, which had been mistakenly depressing Romney's figures? There was little evidence of a significant 'Bradley effect', i.e. where polls overrated Obama because respondents were reluctant to state their opposition to him for fear of seeming racist.[12] But a different possibility was much discussed. It concerned whether polling companies should rebalance samples according to party affiliation. American voters self-identify as one of Democrat, Republican or Independent. If a polling sample were, say, disproportionately composed

---

[11] http://fivethirtyeight.blogs.nytimes.com/2012/11/04/nov-3-romneys-reason-to-play-for-pennsylvania/

[12] http://www.fivethirtyeight.com/2008/08/persistent-myth-of-bradley-effect.html

of Democrats, that would yield a skewed pro-Obama result. In 2012, that was exactly the accusation: polls showed that Romney had a big lead among Independents, and critics charged that Obama came out ahead overall only because the polls were 'over-sampling' Democrats. That is, the proportion of Democrats in samples was charged to be disproportionately high in light of exit polls from previous elections and other considerations.[13]

The key methodological issue is whether party affiliation is a stable population variable that should be adjusted for in the same manner as age or gender, or whether instead it is an unstable variable that is merely an attitude and often just an *effect* of voting preferences. If the latter, then the party of whomever is the more popular candidate may be 'over-sampled' simply because a voter's party self-identification is influenced by their voting intention in a way that their age or gender cannot be. If so, then it is distorting to rebalance for stated party affiliation; but if not so, then it is distorting *not* to. Standard industry practice had been the former, i.e. not to rebalance for stated party affiliation. Predicting correctly who would win the presidency turned critically on whether this practice was correct. Was it?

Again, simply averaging the polls would not help. One piece of evidence gives a flavor of the more detailed kind of analysis required. Across different polls of a particular state, with similar headline figures, there was typically a strong positive correlation between Romney's lead among Independents and the proportion of voters self-identifying as Democrats.[14] The inference from this is that party self-identification is an unstable variable. For various reasons, a given Obama voter might self-identify as Independent in one poll but as Democrat in a second. As a result, in the first poll there are fewer Democrats and Romney's lead among Independents is lower, whereas in the second both are higher – hence the positive correlation. Overall, the important thing from a prediction point of view is therefore whether Obama is leading overall in both polls – as, in swing states, he indeed was.

---

[13] http://www.redstate.com/2012/10/26/why-i-think-obama-is-toast/

[14] http://www.huffingtonpost.com/nick-gourevitch/romney-lead-with-independents_b_2058290.html

5. Failure of the theory-centred approach

The details of poll aggregation show clearly the case-specific nature of its methods. The alternative is to focus instead on 'fundamentals', i.e. on variables that might shed light on election results generally not just case-specifically, such as economic conditions, the perceived extremism of candidates, incumbency, and so forth. There is a literature in political science on election prediction that aims to furnish just such generalizable models.[15] How does it fare?

Conveniently, it too has focused on US presidential elections. The sample size is relatively small, as fewer than 20 elections have good enough data. This creates a danger of overfitting. In response, models typically feature only a small number of variables, most commonly economic ones such as growth in GDP, jobs or real incomes.[16] Sensibly, they are estimated on the basis of one part of the sample and then tested by tracking their predictive performance with respect to the rest of the sample.[17] Even then, there remains a risk of overfitting – if a model predicted the first few out-of-sample elections quite well, will its success continue in future elections? Moreover, even if a model does successfully predict past elections, there is no guarantee the political environment is so stable that the model will remain correct in future too.

---

[15] Influential contributions include Fair (1978), Campbell and Wink (1990), Hibbs (2000), Abramowitz (2008), and Lewis-Beck and Tien (2008). Montgomery et al (2012) averages these and other models to achieve the best forecasting success of all.

[16] Literally thousands of economic variables could plausibly be deemed relevant, not to mention many non-economic ones too.

[17] Of course, there is a long history of debate in philosophy of science about the relative epistemic merits of novel prediction versus retrospective accommodation of the facts. Defenses of the latter typically emphasize how theory may have been developed or tested independently of the particular accommodation, how background knowledge may leave the main epistemic task mere calibration of an agreed functional form, or how a lack of plausible alternative explanations tells in favor of the one that we do find. But none of these defenses applies well to the election prediction case, thus justifying the literature's concentration on prediction.

These caveats noted, it is true that the models do have a little success. On one estimate, the best ones' average error when predicting the incumbent party's share of the vote is between 2 and 3%.[18] But this is not quite as impressive as it might sound: first, for our purposes it is something of a cheat, in that one of the variables in by far the highest weighted model – Abramowitz 2008 – is a polling result, namely presidential approval rating. So the success is not achieved purely by fundamentals. Second, a 2-3% average error corresponds to an average error when estimating the *gap* between the leading two candidates of about 5%. And third, vote shares rarely deviate all that much from 50% anyway, so they are quite an easy target – indeed, another estimate is that economic variables account for only about 30-40% of the *variance* in incumbent party vote share.[19] Overall, the models do not predict individual election results very reliably. On many occasions, they even get wrong the crude fact of which candidate won. For accurate prediction, it is necessary to incorporate the results of opinion polls.

6. Explanation

Why did Obama win? Answering this requires identification of an event's causes.[20] That in turn requires a verified theory or causal model. The problem is that nobody – from either approach – has managed to produce one.

On the polling side, in a trivial sense Obama's victory is 'explained' by the fact that, as revealed by aggregators, on the eve of the election a majority of the electorate were minded to vote for him. But, of course, for most investigative purposes a deeper explanation is required, in particular one that might apply to other elections too. Poll aggregation provides none.[21]

---

[18] See http://www.brendan-nyhan.com/blog/2011/11/a-comparison-of-presidential-forecasting-models.html for discussion and references.

[19] http://fivethirtyeight.blogs.nytimes.com/2011/11/16/a-radical-centrist-view-on-election-forecasting/

[20] Following the literatures under consideration here, I focus on *causal* explanation. I do not mean to rule out the possibility of other forms of explanation.

[21] It is true that some proximate explanations of the election outcome can be *tested* by careful observation of movements in the polls. For instance, the impact of Obama's weak performance in the first presidential

On the fundamentals side, if their models had fared better they would have provided the very explanations that polling aggregation does not. After all, that is precisely the motivation for theory-centred methodology. Thus we might have been able to explain that Obama won because of, say, positive GDP and jobs statistics in the preceding two quarters. Unfortunately, though, the fundamentals models are not predictively accurate.

Can they nevertheless provide us with explanations anyway? The argument would be that they have truly identified relevant causes. It might be postulated, for instance, that GDP or stock market growth does causally impact on voter preferences and thus on election outcomes. True, other causes impact too and so the models do not explain the outcomes fully nor predict them accurately, but that still leaves room for the claim that they explain them 'partially' by correctly identifying *some* of the causes present.[22]

But, alas, even this claim is dubious. First, the different models cite different variables. Abramowitz's, for instance, cites GDP growth, presidential approval rating, and a complex treatment of incumbency; Hibbs's though cites growth in real disposable income and the number of military fatalities abroad. Even among economic variables alone, some models cite GDP, some household incomes, some jobs data, some stock market performance, and so on. There are many different ways to achieve roughly the same limited predictive success, which shakes our faith that any one way has isolated the true causal drivers of election results. Perhaps the small sample size relative to the number of plausible variables makes this problem insoluble.

A second reason for pessimism is that, elsewhere in science, a standard response to predictive failure is to test putative causes in isolation. As it were, at least we achieve predictive success in the isolated test. But unfortunately such experiments are impossible in the case of election predictions. So as well as predictive failure at the level of elections

---

debate, or of Hurricane Sandy in the campaign's final week, can be assessed in this way. But testing an explanation is not the same as providing one.

[22] See Northcott (2013) for more on the relevant sense of partial explanation.

as a whole, the causal factors picked out by the models have not earned their empirical keep by other means either.

The upshot is that we have no warrant for asserting that we have found even some of the causes of election outcomes, and therefore no warrant for claiming even partial explanations. Thus the basic conclusion stands: we have not achieved any explanation of election outcomes, and so the original motivation for turning to fundamentals models is frustrated.

7. Transportability

Are the predictive successes of one election transferrable to another? That is, will a similar polling aggregation strategy work elsewhere? For US presidential elections, it seems that the answer is 'yes' – witness the success of many of the same polling aggregators in 2008. However, it is a different story for other elections, such as US congressional elections or elections in other countries.[23]

The reason is precisely the case-specific nature of polling aggregation – for the best aggregation does not rely only on polls. It must also factor in features such as whether an election is a single national vote or split into many smaller constituencies, whether there are two or many major political parties, whether the voting system is first-past-the-post or proportional representation, one-shot or multi-round, and so on. The implications of a poll for election prediction depend on just such factors. It has also proved predictively profitable to moderate polling results by considering what result should be 'expected', given various local demographic and historical factors. The details of just how to do this are important – and inevitably highly case-specific.

Perhaps even more significantly, the earlier nuances, namely adjudicating state versus national polls and whether polls might be systematically skewed, could also only be resolved by case-specific knowledge. There are many similar examples, such as the

---

[23] The http://fivethirtyeight.blogs.nytimes.com/ predictions of the 2010 US congressional and UK parliamentary elections, for instance, were notably less successful.

extent of regression to the mean to be expected after 'bumps' from party conventions or presidential debates, or if one candidate is 'surprisingly' far ahead at an early stage of the campaign. Such knowledge is crucial, but typically it is transferrable to new elections only imperfectly if at all.

So a serious polling aggregator must build a new election prediction model each time. This lack of transportability is really just the flipside of two facts familiar from above: first, that no one has achieved satisfactory causal explanations – not even the poll aggregators. And second, that predictive success requires case-specific knowledge rather than a search for generalizable causal mechanisms or theoretical underpinnings.

8. Conclusion

How can we make progress, i.e. predict election results even better? It is clear that improving the models of fundamentals is an unpromising route. Rather, progress will be made in the same way as it has been made in the last few years – by doing polling aggregation better. This might involve getting better polling data, analyzing that data better, or understanding better how the implications of that data depend on local peculiarities – in other words, by developing the case-specific, extra-theoretical components of prediction for each application anew. Although transportable explanations are elusive, predictive success need not be. What is clear, though, is that context-free general theory offers neither: it seems there are no short-cuts in social science.

References

Abramowitz, A. (2008). 'It's About Time: Forecasting the 2008 Presidential Election with the Time-for-Change Model', *International Journal of Forecasting* 24: 209-217.

Alexandrova, A. (2008). 'Making models count', *Philosophy of Science* 75, 383-404.

Alexandrova, A. and R. Northcott (2009). 'Progress in economics', in D. Ross and H. Kincaid (eds) *Oxford Handbook of Philosophy of Economics*, Oxford, 306-337.

Brante, T. (2001). 'Consequences of Realism for Sociological Theory-Building', *Journal for the Theory of Social Behaviour* 31, 167–94.

Campbell, J. and K. Wink (1990). 'Trial-Heat Forecasts of the Presidential Vote', *American Politics Quarterly* 18 (3): 251–69.

Cartwright, N. (2007). *Hunting Causes and Using Them*. Cambridge.

Elster, J. (1989). *Nuts and Bolts for the Social Sciences*. Cambridge.

Fair, R. (1978). 'The Effect of Economic Events on Votes for President', *Review of Economics and Statistics* 60, 159-173.

Giddens, A. (1976). *New Rules of Sociological Method: a Positive Critique of interpretative Sociologies*. London: Hutchinson.

Hacking, I. (1995). 'The Looping Effect of Human Kinds', in D. Sperber, D. Premack and A Premack (eds) *Causal Cognition an Interdisciplinary Approach*, Oxford.

Hibbs, D. (2000). 'Bread and Peace Voting in US Presidential Elections', *Public Choice* 104, 149–180.

Lawson, T. (1997). *Economics and Reality*, London: Routledge.

Lewis-Beck, M. and C. Tien (2008). 'The Job of President and the Jobs Model Forecast: Obama for '08?', *PS: Political Science and Politics* 41, 687-90.

Little, D. (1991). *Varieties of Social Explanation*, Boulder, CO: Westview.

Montgomery, J., F. Hollenbach and M. Ward (2012). 'Ensemble Predictions of the 2012 US Presidential Election', *PS: Political Science and Politics* 45, 651-654.

Northcott, R. (2013). 'Degree of explanation', *Synthese* 190, 3087-3105.

Reiss, J. (2008). *Error in Economics: Towards a More Evidence-Based Methodology*. Routledge.

Strevens, M. (2005). 'How are the sciences of complex systems possible?', *Philosophy of Science* 72, 531-556.

Taylor, C. (1971). 'Interpretation and the Sciences of Man', *Review of Metaphysics* 25, 3-51.

All cited 2014 URLs were accessed in November 2014, all earlier ones in August 2013.