This is an author accepted manuscript of an article accepted for publication in Electoral Studies.

Do Party Leadership Contests Forecast British General Elections?

Andreas Erwin Murr

Department of Politics and International Studies, University of Warwick, Coventry, United Kingdom

Correspondence details:

Andreas Murr (corresponding author)
Department of Politics and International Studies
Social Sciences Building
The University of Warwick
Coventry, CV4 7AL
United Kingdom
Telephone: +44 (0)24 765 72959
Email: a.murr@warwick.ac.uk
ORCID: https://orcid.org/0000-0002-9536-0118

Biographical note:
Dr Andreas Murr is Associate Professor in Quantitative Political Science in the Department of Politics and International Relations and a member of the Q-Step Centre at the University of Warwick.

Do Party Leadership Contests Forecast British General Elections?

When assessing election forecasts, two important criteria emerge: their accuracy (precision) and lead time (distance to event). Curiously, in both 2010 and 2015 the most accurate forecasts came from models having the longest lead time—albeit at most 12 months. Can we increase the lead time further, supposing we tolerate a small decrease in accuracy? Here, we develop a model with a lead time of more than 3 years. Our Party Leadership Model relies on the votes of MPs when selecting their party leader. We assess the forecasting quality of our model with both leave-one-out cross-validation and a before-the-fact forecast of the 2019 general election. Compared to both simple forecasting methods and other scientific forecasts, our model emerges as a leading contender. This result suggests that election forecasting may benefit from developing models with longer lead times, and that party leaders may influence election outcomes more than is usually thought.

Keywords: Accuracy; British general elections; Election forecasting; Party leadership contests; Lead time; Leader effects.

1. Introduction

Election forecasts have become a common feature of election commentary in Britain. The increase in both the number and variety of forecasting models is documented well in the pages of *Electoral Studies*. Two special issues of this journal collected forecasts of the 2010 and 2015 general election, with the number of forecasts doubling from six to twelve from one election to the next (Gibson and Lewis-Beck, 2011; Fisher and Lewis-Beck, 2016). The scientific forecasts featured in these special issues represent well the variety of existing forecasting approaches (for a review see Stegmaier and Norpoth, 2017). They include vote intention models (Hanretty et al. 2016), citizen forecasts (Murr, 2016), and political economy models (Lewis-Beck at al. 2016), among others. This increase in both the number and variety of forecasting models is a welcome trend, as forecasting is central to science in general to evaluate both theories and methods (e.g., Lewis-Beck, 2005).

One way to classify the various forecasting models is according to whether they take the long view or the short view (Lewis-Beck and Tien, 2016). These two views differ in how they imagine that an accurate forecast is made. As a result, they also differ in their lead time—how many days before the election they forecast. As Lewis-Beck and Tien explain, short-view models "tend to stress accuracy and to downplay theory and lead". According to this view, accuracy comes from continuously updating the forecast using, for instance, the latest vote intention polls. By contrast, long-view models tend to stress theory and to forecast a good time in advance of the election. According to this view, "[a]ccuracy comes from better theory and an optimal lead time before the election itself".

If we classify the above-mentioned forecasting models of 2010 and 2015 by their lead time, we note that in 2005 the shortest lead time was one month, and the longest was six months (Gibson and Lewis-Beck, 2011: Table 1); in 2015 the shortest was less than a week, and the longest was twelve months (Fisher and Lewis-Beck, 2016: Table 1). Curiously, in both elections the most accurate model—in terms of forecasting the seat share of the incumbent party—was also the model with the longest lead time (Whiteley et al., 2011; Prosser, 2016). This underscores the above point by Lewis-Beck and Tien that models with longer lead time may actually forecast more accurately than models with shorter lead time. This said, we may also prefer models with less accuracy, if they have more lead time than other models (e.g., Lewis-Beck, 1985). Hence an important area of research in forecasting is to examine, as Jennings et al. (2020: 949) put it, "[h]ow much further back can we go, supposing we tolerate a little more error?"

We address this question here by presenting a forecasting model with an average lead time of more than three years, that is two years more than the longest lead time in

2010 and 2015. This forecasting model, the so-called Party Leadership Model, was introduced before the 2015 election (Murr, 2015). At that time, the model had a high accuracy, correctly forecasting *ex post* the winner in 8 out of 10 elections, and since then it correctly forecasted *ex ante* the winner in both 2015 and 2017. However, before the 2019 election, this model only predicted which party leader would become Prime Minister; it didn't predict whether the Prime Minister would have an overall majority or how many seats the Prime Minister would win. Here we first extend this model so that it can forecast these two outcomes of interest, and then assess the success of these extensions.

Before extending the model, the next section presents arguments and evidence suggesting that party leadership contests predict general elections. We then assess the accuracy of the extended model *ex post* in 13 elections. After this, we also assess the accuracy of its first *ex ante* forecast of the 2019 election. We find that the extended Party Leadership Model forecasts the winning party, whether it has an outright majority, and its seats share with high accuracy. We also find that the Party Leadership Model compares favourably to both simple forecasting methods with similar lead time in general, and to other scientific forecasting models with varying lead times in 2019 (Fisher et al., 2019a; Fisher et al., 2019b; Lebo and Fisher, 2019; Murr et al., 2019; Norpoth, 2019; Prosser and Fisher, 2019). Hence the Party Leadership Model represents a parsimonious model that accurately forecasts with long lead time. We conclude by arguing that the success of the Party Leadership Model suggests that election forecasting may benefit from developing models with longer lead time, and that electoral research should take another look at party leader effects on election outcomes.

2. Forecasting British general elections with party leadership contests

Our goal is to forecast British general election outcomes such as which party wins and

what seat share it will gain. The election forecasting literature is flourishing, with steady increases in both the number of forecasters and the diversity of approaches (for a review see Stegmaier and Norpoth, 2017). Here we follow the approach of using leadership contests to forecast general elections. This approach has been pioneered by Norpoth (2001) who used Primary elections to predict US presidential elections. Murr (2015) then adapted the approach to the UK context, using party leadership contests to forecast British parliamentary elections. His adaptation relies on theoretical and empirical work by Sjöblom (1968), Stark (1996), and Quinn (2002) who studied the nature of British party leadership contests. Below we review the theory and evidence of election forecasting with leadership contests in Britain so far, and highlight the contributions of this manuscript.

Helmut Norpoth (2001) pioneered the idea of using information regarding how parties select their candidates to predict how the electorate will vote as a whole. More specifically, his Primary Models uses the main contender's primary performance in New Hampshire to forecasts the US presidential election. Murr (2015: 3–4, 8) then adapted the general idea to the British context with his Party Leadership Model. Because the political system of the two countries differ, the theory and data underlying the two models necessarily differ too. The theory and evidence of the party leadership model will be discussed in more detail below, but in short the differences between the two models are as follows. Norpoth forecasts presidential elections; we forecast parliamentary elections. Norpoth relies on both party members and independents voting in the New Hampshire primary; we rely on MPs voting in leadership contests. Norpoth argues that primary performance measures the public's satisfaction with government; we argue that performance in leadership contests measure MPs' expectations about the

party winning the election. Despite these differences, both models suggest that leadership contests forecast general elections across political systems.

More specifically, our claim is that party leadership contents forecast general elections in Britain. We begin the reasoning underlying this claim by describing the strategic goals of political parties. According to Sjöblom (1968) major parties in parliamentary systems pursue three strategic goals: party unity, electoral victory, and policy implementation. These goals are listed in decreasing order of importance with higher-order goals necessary, but not sufficient for lower-order goals. In other words, no policy implementation without electoral victory; no electoral victory without party unity. One implication of this view is that a party chooses as a leader the candidate who allows the party to achieve the highest-order goal. Testing this theory in British party leadership contests, Stark (1996:126) and Quinn (2012: 13) have found that most British leadership contests have focused on achieving electoral victory.

Next, we argue that a party's Members of Parliament (MPs) have the means, motive, and opportunity to select a party leader with the highest electoral appeal. First, the MPs know the leadership candidates like nobody else does ("means"). The candidates are MPs themselves and so selectors and candidates observe each other on a regular basis. Second, the future of MPs depends on selecting a party leader with the highest chance of winning the election ("motive"). According to King (2002: 4–5) leaders can influence the election outcome either indirectly by "changing [their] party's ideology or modernizing its image" or directly via their "personalities and personal characteristics". Finally, since 1963 both the Conservatives and Labour have allowed MPs to formally vote in party leadership contests ("opportunity"). In sum, the relative performance of elected leaders in leadership contests should indicate their chances of

winning the general election. It follows that party leadership contests should forecast general elections.

The empirical support for this claim is strong, as demonstrated by the Party Leadership Model (Murr 2015). For each party leadership contest, it calculates the difference in vote share between elected leader and main contender in the nomination process or final ballot among MPs (see Table 1). The model then picks the leader with the larger difference as the predicted winner of the general election (see Table 2). Since the model was introduced it correctly predicted *ex ante* the re-elections of David Cameron in 2015 and of Theresa May in 2017. In addition, when applied *ex post*, that is in retrospect, it correctly predicted 8 out of 11 general elections between 1966 and 2010. This is a remarkable accuracy given that the model can forecast as soon as we know who leads the two main parties into a general election. In the past this has been more than three years in advance on average (see Table 3). (The data and code to replicate all results have been submitted as supplemental files to the journal.)

Table 1: Performance of Conservative and Labour party leaders in their party leadership election among MPs in per cent.

General election	Party leadership election	Ballot	Name		Vote share		Performance
			Elected leader	Main contender	Elected leader (v_l)	Main contender (v_c)	(v_l - v_c)
Conservatives							
1966–1974	1965	1st	Edward Heath	Reginald Maudling	50.3	44.6	5.7
1979–1987	1975	2nd	Margaret Thatcher	William Whitelaw	52.9	28.6	24.3
1992	1990	2nd	John Major	Michael Heseltine	49.7	35.2	14.5
1997	1995	1st	John Major	John Redwood	66.3	27.1	39.2
2001	1997	3rd	William Hague	Kenneth Clarke	56.8	43.2	13.6
2005	2003	—	Michael Howard	—	—	—	—
2010–2015	2005	2nd	David Cameron	David Davis	45.5	28.8	16.7
2017	2016	2nd	Theresa May	Andrea Leadsom	60.5	25.5	35.0
2019	2019	5th	Boris Johnson	Jeremy Hunt	51.1	24.6	26.5
Labour							
1966–1974	1963	2nd	Harold Wilson	George Brown	58.3	41.7	16.6
1979	1976	3rd	James Callaghan	Michael Foot	56.2	43.8	12.4
1983	1980	2nd	Michael Foot	Denis Healey	51.9	48.1	3.8
1987	1983	1st	Neil Kinnock	Roy Hattersley	49.3	26.1	23.2
1992	1988	1st	Neil Kinnock	Tony Benn	82.8	17.2	65.6
1997–2005	1994	1st	Tony Blair	John Prescott	60.5	19.9	40.6
2010	2007	—	Gordon Brown	John McDonnell	88.2	8.2	80.0
2015	2010	4th	Ed Miliband	David Miliband	46.6	53.4	-6.8
2017–2019	2015	—	Jeremy Corbyn	Andy Burnham	15.5	29..3	-13.8

Note: This table only includes party leadership elections that selected a party leader who stood in a General Election. Michael Howard (Conservative Party in 2003) faced no contender in his party leadership election. For him a performance measure is unavailable. Gordon Brown (Labour Party in 2007) faced a contender (John McDonnell) in the nomination rounds, but McDonnell did not get the necessary number of nominations to secure a place on the ballot, so Brown was the only successfully nominated candidate. *Source*: Quinn (2012) and own calculations.

Table 2: The Party Leadership Model correctly predicts 10 out of 13 past elections (1966–2017) when data is sufficiently available to make a forecast.

General election	Incumbent	Performance of party leaders (v_l - v_c)		Prediction	Winner
		Conservative	Labour		
1966	Labour	5.7	16.6	Labour	Labour
1970	Labour	5.7	16.6	Labour	Conservative
1974 (Feb)	Conservative	5.7	16.6	Labour	Labour
1974 (Oct)	Labour	5.7	16.6	Labour	Labour
1979	Labour	24.3	12.4	Conservative	Conservative
1983	Conservative	24.3	3.8	Conservative	Conservative
1987	Conservative	24.3	23.2	Conservative	Conservative
1992	Conservative	14.5	65.6	Labour	Conservative
1997	Conservative	39.2	40.6	Labour	Labour
2001	Labour	13.6	40.6	Labour	Labour
2005	Labour	—	40.6	—	Labour
2010	Labour	16.7	80.0	Labour	Conservative
2015	Conservative	16.7	-6.8	Conservative	Conservative
2017	Conservative	35.0	-13.8	Conservative	Conservative
2019	Conservative	26.5	-13.8	Conservative	Conservative

Note: The leaders of the Conservatives faced no contender in the party leadership elections relevant for the 2005 General Election. For this General Election a forecast is unavailable.

Table 3: Forecasting lead of the Party Leadership Model. The lead equals the number of days between the dates of the last party leadership election and of the General Election.

General Election	Party leadership election		Forecast	Lead
	Conservatives	Labour		
31/03/66	28/07/65	14/02/63	28/07/65	246 days
18/06/70	28/07/65	14/02/63	28/07/65	1786 days
28/02/74	28/07/65	14/02/63	28/07/65	3137 days
10/10/74	28/07/65	14/02/63	28/07/65	3361 days
03/05/79	11/02/75	03/04/76	03/04/76	1125 days
09/06/83	11/02/75	10/11/80	10/11/80	941 days
11/06/87	11/02/75	02/10/83	02/10/83	1348 days
09/04/92	27/11/90	02/10/88	27/11/90	499 days
01/05/97	04/07/95	21/07/94	04/07/95	667 days
07/06/01	19/06/97	21/07/94	19/06/97	1449 days
05/05/05	06/11/03	21/07/94	—	—
06/05/10	06/12/05	24/06/07	24/06/07	1047 days

07/05/15	06/12/05	25/09/10	25/09/10	1685 days
08/06/17	11/07/16	12/09/15	11/07/16	332 days
12/12/19	22/07/19	12/09/15	22/07/19	143 days
				Mean = 1269 days

Source: Quinn (2012) and own calculations.

Of course, to evaluate whether this forecasting accuracy is relatively good or bad, we need a benchmark for comparison. Campbell (2008: 262) proposed several such benchmarks. They vary in difficulty of surpassing them: from easy ("random guess") to demanding ("the polls conducted just prior to the election"). Although Campbell developed these benchmarks for evaluating forecasts of vote shares in US presidential elections, these can be adapted in a straightforward manner for evaluating forecasts of who wins in British general elections. The Party Leadership Model clears the first benchmark: as indicated above, it correctly predicted 10 out of 13 past elections, which is better than chance. It also clears the second benchmark (Murr 2015: 5f): when it is reasonable to compare the Party Leadership Model to polls in the month before the election, the forecasting accuracy of the Party Leadership Model is 86% compared to 43% for the polls. Even with an optimal lead time the polls achieve an accuracy of only 71%, which is still below the Party Leadership Model. Hence, at least when forecasting who wins, the Party Leadership Model does better than both chance and vote intention polls.

In sum, these findings support the idea that party leadership contests forecast who wins the general elections, and that they do so better than public opinion polls. Our overall aim is to see if these findings generalise to other forecasting targets and to other election forecasting models. Our first goal is to see if party leadership contests also forecasts whether the winner will have an outright majority, and what seat share the Conservatives and Labour will win. Our second goal is to see if the forecasting

accuracy of the Party Leadership Model compares favourably to both simple forecasting methods and other scientific forecasting models. To achieve both goals, we made public a forecast with the Party Leadership Model of the 2019 general election on the 10th of December, two days before the election.[1]

3. Forecasting who wins and whether the winner has an outright majority

As indicated above, the Party Leadership Model in its initial form forecasted whether the incumbent party will be re-elected or not. Here the incumbent party is the party with the most seats. With hung parliaments occurring more regularly in recent times, and having consequences for stability of governments and their ability to govern, it is of course important to also be able to forecast whether hung parliaments occur, or at least how likely it is that they do. In this section, we extend the Party Leadership Model to do just that. We then test the hypothesis that having the more popular leader makes it more likely for the incumbent party to win an outright majority. Finally, we assess the accuracy of this extension via cross-validation and an *ex ante* forecast of the 2019 election.

To extend the Party Leadership Model this way, we only need to extend its outcome variable from two possible values (0 = *incumbent party not re-elected*; 1 = *incumbent party re-elected*) to four possible values (1 = *outright majority for opposition party*, 2 = *hung parliament led by opposition party*, 3 = *hung parliament led by incumbent party*, 4 = *outright majority for incumbent party*). We can then test whether

[1] To allow for a transparent test of the predictions, the forecast was made public before the result of the election was known. The blog electionsetc.com featured a post about the forecast that went online on the 10th of December. A link to the post has been submitted to the editors and will be referenced, if the manuscript is accepted.

the binary predictor (0 = *opposition-party leader more popular*, 1 = *incumbent-party leader more popular*) correlates with the outcome variable.

Table 4 shows the joint frequency distribution of the two variables, with the outcome in the rows and the predictor in the columns. The cells display the number of observations in parentheses as well as the posterior probability of the outcome given the value of the predictor.[2] For instance, the incumbent party had the more popular leader in 9 elections; it then won an outright majority in 6 of them. This means that if the incumbent party has the more popular leader, its chance of winning an outright majority is 59.1 per cent. (One may wonder why our answer differs from 6 / 9 = 66.6 per cent. The reason is that we include one half of an observation as a prior in each cell. When done in this way, the answer is 6.5 / 11 = 59.1 per cent. See Footnote 2 for more details.)

Table 4: Posterior probabilities of general election outcomes by relative leader popularity among MPs.

	More popular leader	
	Opposition party	Incumbent party
Outright majority for incumbent	25.0 (1)	59.1 (6)
Incumbent-led hung parliament	8.3 (0)	13.6 (1)
Opposition-led hung parliament	25.0 (1)	13.6 (1)
Outright majority for opposition	41.7 (2)	13.6 (1)
	100.0 (4)	100.0 (9)

Note: Number of observations in parentheses.

[2] Formally, we have an outcome y which takes on value k=1,2,3,4. Our predictor x can take on two possible values, 0 and 1. We denote the probability that $y=k$ conditional on x as $p_k(x)$. The prior distribution for $p_k(x)$ conditional on x is Dirichlet with prior sample size of .5 for each category. We denote the frequency of outcome $y=k$ conditional on x as $n_k(x)$. The posterior distribution for $p_k(x)$ is then also Dirichlet, with posterior sample size $.5+ n_k(x)$. The posterior mean of $p_k(x)$ is then $[.5+ n_k(x)] / [2 + n_1(x) + n_2(x) + n_3(x) + n_4(x)]$. We use MCMC simulation to compute the probability that $p_4(1) + p_3(1) > p_4(0) + p_3(0)$ and that $p_4(1) > p_4(0)$. See Murr (2015) for a similar analysis where K=2 and Gelman et al. (2014: 69f, 578f) for more details on Bayesian analysis with the Dirichlet distribution.

Before using this table to test the main hypothesis of interest—if the incumbent party has the more popular leader, its chances of winning an outright majority are higher—we first update the test of a simpler, more general hypothesis examined first by Murr (2015). This hypothesis reads as follows: if the incumbent party has the more popular leader, its chances of re-election are higher. Murr (2015) estimated the probability of this hypothesis being true as 95 per cent. Here, we bring the analysis up to date by including the Brown, Cameron, and May elections.

An initial glance at Table 4 suggests that the empirical support for this hypothesis remains strong. When the incumbent-party leader was more popular, the incumbent was re-elected 7 out of 9 times, but when its leader was less popular, the incumbent party was re-elected only 1 out of 4 times. Indeed, a Bayesian analysis puts the probability of the hypothesis being true at 95 per cent (see Footnote 2). That is, we have strong evidence that if the incumbent party has the more popular leader, its chances of re-election are higher (see also Murr 2015).

Next, we examine the new hypothesis that if the incumbent party has the more popular leader, its chances of winning an outright majority are higher. An initial glance at Table 4 suggests that this is also the case. When the incumbent-party leader was more popular, the incumbent won an outright majority 6 out of 9 times, but when its leader was less popular, the incumbent party won an outright majority only 1 out of 4 times. Indeed, a Bayesian analysis puts the probability of the hypothesis being true at 93 per cent (see Footnote 2). That is, we have also strong evidence that if the incumbent party has the more popular leader, its chances of winning an outright majority are higher.

This finding implies that the Party Leadership Model should forecast whether there will be a hung parliament or an outright majority. Before testing this implication

with a before-the-fact forecast of the 2019 general election, we first test it with an after-the-fact test—a leave-one-out cross-validation. We leave out one election at a time from the data set that we fit an ordinal regression model to—both the outcome and predictor are the same as before. We then predict the left-out election based on the fitted model. Finally, we compare the predicted with the actual result. We use two measures of accuracy: whether the prediction was correct ('hit') and what the predicted probability of the actual outcome was ('likelihood'). Consequently, we get 13 out-of-sample measures each, which we average to get both a 'hit rate' and an 'average likelihood'.

How does our model perform in this cross-validation test? As a benchmark we consider a 'random guess': it would put about 25 per cent of likelihood on the actual event and would hit the actual target about 25 per cent of the time. By contrast, we find that our model puts about 36.9 per cent of likelihood on the actual event and hits the target about 61.5 per cent of the time (no table shown). This means that our model predicts much better than chance.

We use this cross-validation also to justify our binary measure of party leadership compared to a continuous one (the difference in party leadership). One may think that the continuous measure should predict better than the binary one. However, this is not the case. In an ordinal regression model, we compared the performance of the binary and the continuous predictors in a leave-one-out cross-validation on the same two measures as above: the proportion of correctly predicted outcomes ('hit rate') and the average predicted probability of the actual outcome ('average likelihood'). Compared to the binary predictor, the continuous one decreased the hit rate from 61.5 to 46.2 per cent, and decreased the average likelihood from 36.9 to 29.5 per cent (no table

shown). This means that the simple binary measure predicts better than the continuous measure. (See also results in Section 3.)

Finally, we move on to evaluate the first *ex ante* forecast of this extended model. What did the model forecast for the 2019 election? We made public this forecast on the 10th of December, two days before the election (see Footnote 1). In July 2019, the Conservatives elected Boris Johnson as their party leader. In the 5th and final ballot among MPs, he received 51.1 per cent of the votes, whereas his main contender, Jeremy Hunt, received 24.6 per cent. Later on, party members selected Boris Johnson as their leader. Accordingly, Johnson's popularity among MPs was 51.1 – 24.6 = 26.5 per cent. In September 2015, Labour elected Jeremy Corbyn as their party leader. In the nomination process among MPs, he received 15.5 per cent of the votes, whereas his main contender, Andy Burnham received 29.3 per cent. Later on, party members and registered or affiliated supporters selected Jeremy Corbyn as leader of the Labour Party. Accordingly, Corbyn's popularity among MPs was 15.5 – 29.3 = –13.8 per cent. Because the performance among MPs of Boris Johnson (26.5 per cent) was better than of Jeremy Corbyn (–13.8 per cent), the Party Leadership Model predicted a re-election of Boris Johnson. The most likely outcome was him winning an outright majority, with 59.1 per cent probability. The chances of re-election with or without outright majority were 59.1 + 13.6 = 72.7 per cent. In other words, the Party Leadership Model correctly forecasted for 2019 the re-election of the Conservative government with an outright majority.

4. Forecasting seat shares

4.1 Comparing forecasting models

Next, we take on the task of forecasting seat shares with the Party Leadership Model.

To do so, we build a simple regression model of the seat share of the incumbent or opposition party with a single predictor: whether its leader was more popular than the other. We then assess both its after-the-fact forecasts of 13 elections between 1966 and 2017 and its before-the-fact forecast of the 2019 election. This assessment also includes comparisons with both simple forecasting methods and other scientific election forecasts.

When comparing forecasting models, Lewis-Beck (2005: 151–154) suggests combining four components—accuracy, parsimony, reproducibility, and lead—into an overall quality index according to a formula (see also Lewis-Beck 1985). Because we consider all models discussed below equally parsimonious and reproducible, we only need to know that the quality formula treats accuracy and lead as multiplicative. This means, for instance, that if two models have similar lead time, then the more accurate model is of higher quality. This means also, to give another example, that if the first model is less accurate, but has sufficiently more lead time than the second one, then the first model can still be of similar or even higher quality than the second model.

Below we first compare the Party Leadership Model with simple forecasting methods of similar lead time. Because the models have similar lead time, their overall quality index equals their accuracy. Then, we compare the Party Leadership Model with other scientific models of varying lead times. Because the models have different lead times, their overall quality index considers both accuracy and lead time. Both comparisons will tell us about the forecasting quality of the Party Leadership Model.

4.2 Comparison with simple forecasting methods of similar lead time

Does the Party Leadership Model forecast more accurately than simple forecasting methods? We consider three such simple forecasting methods as benchmarks. Two of these simple forecasting methods are commonly used as benchmarks in the forecasting

literature. For instance, in time-series analysis they are called the "average method" and the "naïve method" (e.g., Hyndman and Athanasopoulus, 2021: Section 5.2); in meteorology they are known as "climatology" and "persistence" (e.g., Murphy, 1992). These two benchmarks are the historical average (the History Model) and the previous value (the Persistence Model). The forecast of the Historical Model equals the historical seat share average of incumbent or opposition parties; the forecast of the Persistence Model equals the previous seat share.

We add a third benchmark forecast that contains the History Model and the Persistence Model as special cases: this third forecast comes from a regression model of the current seat share that uses the previous seat share as its only predictor (the Autoregressive Model). As can be easily seen, if the coefficient of the previous seat share is zero, then the Autoregressive Model becomes the History Model; if the intercept is zero and the coefficient of the previous seat share is one, then the Autoregressive Model becomes the Persistency Model. Hence, this regression model raises the bar compared to the other two benchmarks—it makes fewer assumptions and uses data to approximate the optimal combination of the historical average and the previous seat share. Like the other two benchmarks, this one can be assessed on the same data (past and future) as the Party Leadership Model. Similarly, all three benchmarks have similar lead times to the Party Leadership Model. For these reasons, we think that the History, Persistence, and Autoregressive models are useful points of reference for evaluating election forecasting models in general and the Party Leadership model in particular.

Table 5 shows the estimates for each regression model based on the 13 elections described above. Table 5 displays six regression models in total, one for the incumbent party and one for the opposition party for each of three forecasting models. (The

Persistence Model requires no estimation.) All three regression models show an incumbency advantage: the intercepts for the incumbent party are higher than for the opposition party. The History Model shows that on average the incumbent party wins a higher seat share than the opposition party. The Persistence Model demonstrates that this holds true even if both parties had only slightly different previous seat shares. Finally, the Party Leadership Model confirms that this also holds true even if the incumbent-party leader is less popular than the opposition-party leader. This said, the party with the more popular leader is predicted to gain a higher seat share than the party with the less popular leader. If the incumbent party has the more popular leader, its expected seat share is 11.3 points higher than if it hasn't. For the opposition party this expected difference is 12.4, and so is comparable in size.

Table 5: Regression models of seat share (%) for the incumbent and opposition party for 13 elections (1966–2017), together with an *ex ante* forecast for 2019.

		History Model	*Persistence Model*	*Autoregressive Model*	*Party Leadership Model*
Incumbent party	Intercept	49.2 (2.8)		69.9 (14.6)	41.4 (4.4)
	Previous seat share		1.0	-0.4 (0.3)	
	Leader performed better				11.3 (5.3)
	RMSE (out-of-sample)	10.5	15.7	10.3	10.0
	Coverage of 95% PI (out-of-sample)	92%	92%	92%	92%
	2019 forecast	49.2	48.8	49.1	52.7
	95% PI	[26.5,71.9]	[14.9,82.6]	[27.1,71.1]	[32.1,73.2]
Opposition party	Intercept	42.9 (2.8)		61.9 (13.1)	39.1 (2.8)
	Previous seat share		1.0	-0.4 (0.3)	
	Leader performed better				12.4 (5.1)
	RMSE (out-of-sample)	10.5	15.8	10.6	9.3
	Coverage of 95% PI (out-of-sample)	92%	92%	92%	100%
	2019 forecast	42.9	40.3	44.5	39.1
	95% PI	[20.1,65.6]	[6.1,74.5]	[22.5,66.6]	[19.4,58.7]

Note: Standard errors in parentheses; SE = standard error of the regression; RMSE = root mean squared error; *n* = number of observations; PI = Prediction interval.

We evaluate the forecasting accuracy of the models with two out-of-sample test. Both tests assess how well the models forecast new data that were not used when fitting the models. Out-of-sample tests assess forecasting accuracy much better than in-sample fit statistics; the reason is that maximising fit statistics can lead to over-fitting the model to the data, whereas out-of-sample tests avoid this issue. The first test assesses the models using leave-one-out cross-validation. The second test consists of an *ex ante* forecast of the 2019 election.

First, we follow the leave-one-out cross-validation strategy. We leave out one election at a time from the data set that we fit the regression models to. We then predict the left-out election based on the fitted models. Finally, we compare the predicted with the actual result. Consequently, we get 13 out-of-sample prediction errors, which we square and then take the root mean of. In short, our comparison uses the root mean squared error (RMSE) as computed by leave-one-out cross-validation as its first measure of forecasting accuracy.

Table 5 shows the root mean squared error (RMSE) in the leave-one-out cross-validation exercise for each model. The most accurate forecasting model for both incumbent and opposition parties is the Party Leadership Model. For the incumbent party the most accurate model is the Party Leadership Model (10.0), followed by the Autoregressive Model (10.3). For the opposition party the most accurate model is again the Party Leadership Model (9.3), followed this time by the History Model (10.5). In other words, relative to the best benchmark, the Party Leadership Model increases the forecasting accuracy by about 1 – (10.0 / 10.3) = 3 per cent for the incumbent party and 1 – (9.3 / 10.5) = 11 per cent for the opposition party.

One may think that more complex versions of the Party Leadership Model would predict better. According to this view, adding seat shares to the Party Leadership

Model or replacing its binary popularity measure with a continuous one would increase out-of-sample accuracy. However, this is not the case. Adding seat shares as a predictor to the Party Leadership Model increases the out-of-sample RMSE from 10.0 to 10.2 for the incumbent party, and from 9.3 to 9.4 for the opposition party. Similarly, replacing the binary measure with a continuous measure increases the out-of-sample RMSE. Replacing the binary measure with the difference in popularity between the two leaders increases the out-of-sample RMSE from 10.0 to 10.6 for the incumbent party and from 9.3 to 11.7 for the opposition party; likewise, replacing the binary measure with the leader popularity increases the out-of-sample RMSE from 10.0 to 11.2 for the incumbent party, and from 9.3 to 11.0 for the opposition party (no table shown). This means that the simple binary measure of whether the leader is more popular predicts best.[3]

Table 5 also shows the coverage of the 95% prediction intervals in the leave-one-out cross-validation exercises for each regression model. Whereas confidence intervals contain a fixed quantity such as a parameter in frequentist statistics with a specified probability, prediction intervals contain a random variable such as a future observation with a specified probability (Meeker et al. 2017). Hence prediction intervals are the appropriate intervals to quantify predictive uncertainty. The 95% prediction intervals should contain the actual out-of-sample results in 95% of the cases.

[3] In general, dichotomising a continuous variable can be beneficial when, for instance, there are outliers (DeCoster et al. 2009) or the continuous variable has measurement error (Gustafson and Le 2002). In our case, for instance, the Labour loss in 2010 despite the unusually high popularity measure of Brown could be considered an outlier. Further, leadership contests could be considered a noisy measure of electability: for instance, it is likely that reruns of the same contest would produce slightly different votes due to undecided MPs. As a result, we observe the better performance of the binary predictor compared to the continuous one.

With 13 observations we cannot, of course, reach coverage of exactly 95%, which would mean that 12.4 intervals cover the actual out-of-sample result. Instead, we would be content if 12 or 13 out of the 13 intervals would do so, that is coverage of 92 or 100%, respectively. Table 5 shows that the coverage of the 95% prediction intervals is indeed 92% in all models except for one where it is 100%. This means that despite the small sample size the prediction intervals do well in capturing the uncertainty of the forecasts.

While informative, the leave-one-out cross-validation is not a genuine forecast—it does not forecast an election with yet unknown outcome. The 2019 British general election offered the first opportunity to assess the seat share forecasts of the Party Leadership Model, and to compare it with both simple forecasting methods and other scientific forecasting models. Below, we assess how well the models did in their genuine forecast of the 2019 election. As mentioned before, these forecasts were made public on the 10th of December, two days before the election date (see Footnote 1).

How did the Party Leadership Model compare to the three simple forecasting methods? Table 5 displays the forecasts for the incumbent and main opposition parties in the last two rows of their respective panels. Because the Conservatives were the incumbent party with the more popular leader, the Party Leadership Model forecasted that they would gain 52.7 per cent of seats, while Labour, the main opposition party, would gain 39.1 per cent of seats. Translated into seat numbers, it forecasted that the Conservatives would win 342 seats (they won 365), and that Labour would win 254 seats (they won 202).

This forecast means that the Party Leadership Model was again more accurate than the History, Persistence, and Autoregressive Models. For the seat share of the Tories, the RMSE was 7.4 per cent for the Persistence Model and 7.0 per cent for both

the History and the Autoregressive Model. By contrast, it was only 3.5 per cent for the Party Leadership Model. Similarly, for the seat share of Labour, the RMSE was 13.4 per cent for the Autoregressive Model, 11.8 per cent for the History Model, and 9.2 per cent for the Persistence Model. By contrast, it was only 8.0 per cent for the Party Leadership Model. In sum, the Party Leadership Model continued to make accurate forecasts about British general elections with a long time, at least when compared with simple forecasting methods. How did the Party Leadership Model compare to other scientific election forecasting model in 2019?

4.3 Comparison with other scientific election forecasting models of varying lead times

The 2019 election offers an opportunity to evaluate the overall quality of the several scientific election forecasting models, including the Party Leadership Model. As indicated above, we judge these models by their overall quality index as proposed by Lewis-Beck (2005). We think it fair to say that all these models are equally parsimonious and reproducible; then, their overall quality index equals their accuracy (rated 0–3) multiplied by their lead time (0-3) divided by the highest possibly score for the numerator, 9. As a result, the quality index ranges from 0 to 1.[4]

The first column of Table 6 lists the different models. In addition to the Party Leadership Model, the list includes both structural models such as the local elections model (Prosser, 2016), the PM and Pendulum model (Lebo and Norpoth, 2007), and the citizen forecasting model (Murr et al., 2021) as well as the PSA Expert Survey (Fisher et al., 2019a), polling averages with uniform national swing, and betting markets (Fisher

[4] A forecast could score 0 for accuracy if, for instance, it predicts the wrong party to win; and it could score 0 for lead if, for instance, it makes its "forecast" after the election (Lewis-Beck, 2005: 154).

et al., 2019b). The next two columns note the models' forecasted seat number for the Conservatives and Labour, respectively. The fourth column reports the accuracy score.

Table 6: Forecasts of the 2019 British general elections.

Lead time	Seats		Quality		
Model	Con	Lab	Accuracy	Lead	Overall
4-6 months					
Local elections: Prosser and Fisher	329	231	1	3	0.4
Party Leadership Model	342	254	2	3	0.7
2 months					
PM and Pendulum: Lebo and Fisher	311	268	1	2	0.3
PM and Pendulum: Norpoth	352	228	3	2	0.7
Citizens: Murr, Stegmaier and Lewis-Beck	360	190	3	2	0.7
1 week					
PSA Expert Survey	324	233	1	1	0.1
UNS from Poll Average	345	224	2	1	0.3
Betting markets average	346	221	2	1	0.3
Actual	*365*	*202*			

Note: Local elections (Prosser, 2016; Prosser and Fisher, 2019); Party Leadership Model (Murr, 2015; Murr, 2019); PM and Pendulum (Lebo and Fisher, 2019; Lebo and Norpoth, 2007; Norpoth, 2019); Citizens (Murr et al., 2019; Murr et al., 2021); PSA Expert Survey (Fisher et al. 2019b); UNS from Poll Average and Betting markets average (Fisher 2019a); Accuracy = the score (0–3) on accuracy; Lead = the score (0–3) on lead time; Overall = the quality index calculated from this formula: (Lead x Accuracy) / 9.

All models correctly predicted the winner (the party with the most seats), so they all get one point for accuracy at least. We see, however, that the models differ greatly in their forecasted seat number of the Conservatives. The three worst models—the PM and Pendulum model by Lebo and Fisher, the local elections model, and the PSA Expert Survey—forecasted either a hung parliament or a narrow majority for the Conservatives. Hence they don't get any additional accuracy points. The two best models—the citizen forecasting model followed by the PM and Pendulum model by Norpoth—came within 13 seats of the actual seat number. This means the highest accuracy score of 3. The Party Leadership Model, together with the polls and betting markets, fall between these two extremes: they correctly forecasted a sizeable majority

for the Conservatives, but underestimated its true size by about 20 seats. As a result, they score 2 points for accuracy.

Next, we consider lead time (penultimate column). All models forecasted before the election, so they all get one point for lead at least. We see, however, that the models differ greatly in their lead time. Three models have a lead time of one week or less (1 point in total); three models of about two months (2 points); and two models, including the Party Leadership Model, of four to six months (3 points). Of course, readers may judge both accuracy and lead scores for themselves and decide whether they agree with ours.

How do these two criteria combine to the overall quality index? These quality scores are reported in the last column. The three best models with a score of 0.7 are the citizen forecasting, the PM and Pendulum (Norpoth), and the Party Leadership models. By way of illustration, the Party Leadership Models' accuracy score is 2 and its lead score is 3, yielding an overall quality index of (2 x 3) / 9 = 0.7. The reason the Party Leadership Model features at the top is that despite its lower accuracy compared to the other two models, it has a longer lead time. This performance, together with the fact that its lead time is usually years and not months, suggests that the Party Leadership Model is emerging as a leading model in the forecasting enterprise.

5. Discussion and Conclusion

Both theory and evidence indicate that most party leadership contests in Britain focus on achieving electoral victory. Parties tend to select the leader who maximises its chance of winning the next general election (e.g., Quinn, 2002). If this is true, and if MPs are good at judging leadership candidates, then leadership contests should forecast general elections. Here, we tested this argument in a new way. Murr (2015) already showed that leadership contests forecast which party wins the general election. We

extended this work by showing that these contests also forecast whether the winning party has an outright majority and how many seats the winning party will win. The so-called Party Leadership Model forecasts election outcomes well with a long lead time. Its average lead time of about three years is two years more than any other scientific election forecasting model from 2010 to 2019. These results suggest that election forecasting may benefit from developing models with longer lead time. In addition, these results strongly support the view parties want to achieve electoral victory when selecting a new leader.

Most citizens, journalists, and politicians would readily agree to this; indeed, they commonly believe that leaders matter for elections. By contrast, most electoral researchers tend to be more sceptical. While many have found that a voter's evaluations of party leaders may affect his or her vote choice, they have also noted that this effect tends to cancel out when aggregated to the electorate as a whole. In other words, citizens' leadership evaluations usually have little effect on the vote share of a party, and they rarely influence which party wins the election (e.g., Graetz and McAllister, 1987; Stewart and Clarke, 1992; Crewe and Bartle, 2002). We think the difference between these and our findings can be easily resolved by remembering the distinction mentioned above between direct and indirect leadership effects (King 2002: 4–6).

Most electoral researchers have focused on direct leadership effects. They have measured citizens' evaluations of a leader's personality and personal characteristics such as physical appearance, honesty, and intelligence. By contrast, we have focused on the total leadership effect. We argued that MPs' evaluations of a leadership candidate include both direct leadership effects (via personality) and indirect leadership effects (via changes to their party). Our findings suggest, therefore, that electoral

research may find stronger leadership effects by broadening its focus to also consider how leaders change their party's organisation, image, and ideology.

Acknowledgements

I am very grateful to Ericka Rascón Ramirez and Steve Fisher for useful discussions, comments, and encouragement. I am also thankful for the feedback and guidance received from Kaat Smeets and the anonymous reviewers. All remaining errors are my own. A replication archive will be made available at Harvard Dataverse.

References

Bartle, J. & Crewe, I. (2002), 'The Impact of Party Leaders in Britain: Strong Assumptions, Weak Evidence', in: King, A. (ed.), Leaders' Personalities and the Outcomes of Democratic Elections, Oxford: Oxford University Press.

Campbell, J. E. (2008), 'Evaluating U.S. presidential election forecasts and forecasting equations', International Journal of Forecasting 24, 259–271.

DeCoster, J., Iselin, A.-M. R., & Galluci, M. (2009), 'A Conceptual and Empirical Examination of Justifications for Dichotomization', Psychological Methods 14(4), 349–366.

Fisher, S., Kenny, J., & Shorrocks, R. (2019a). 'Final combined forecast for the 2019 general election', published 12 December 2019, viewed 25 January 2021, https://electionsetc.com/2019/12/12/final-combined-forecast-for-the-2019-general-election/

Fisher, S., Kirby, M., & Greenwood, J. (2019b). 'PSA expert predictions report: 2019 general election', published 9 December 2019, viewed 25 January 2021, https://www.psa.ac.uk/psa/news/psa-expert-predictions-report-2019-general-election

Fisher, S., & Lewis-Beck, M. (2016). 'Forecasting the 2015 British general election: The 1992 debacle all over again?', Electoral Studies 41, 225–229.

Graetz, B., & McAllister, I. (1987). 'Party Leaders and Election Outcomes in Britain, 1974–1983', Comparative Political Studies 19, 484–507.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014), Bayesian Data Analysis, third edn, Boca Raton, Florida: CRC Press.

Gibson, R., & Lewis-Beck, M. (2011) 'Methodologies of election forecasting: Calling the 2010 UK "hung parliament"', Electoral Studies 30, 247–249.

Gustafson, P., & Le, N. D. (2002). 'Comparing the Effects of Continuous and Discrete Covariate Mismeasurement, with Emphasis on Dichotomization of Mismeausured Predictors', Biometrics 58(4), 878–887.

Hanretty, C., Lauderdale, B., & Vivyan, N. (2016). 'Combining national and constituency polling for forecasting', Electoral Studies 41, 239–243.

Hyndman, R. J., &Athanasopoulos, G. (2021), Forecasting: Principles and Practice, third edn., OTexts: Melbourne, Australia, viewed on 17 February 2021, https://OTexts.com/fpp3

Jennings, W., Lewis-Beck, M., & Wlezien, C. (2020), 'Election forecasting: Too far out?', International Journal of Election Forecasting 36: 949–962.

King, A. (2002), 'Do Leaders' Personalities Really Matter?', in: King, A. (ed.), Leaders' Personalities and the Outcomes of Democratic Elections, Oxford: Oxford University Press.

Lebo, M., & Fisher, S. (2019), 'Forecasting the 2019 General Election using the PM and the Pendulum model', published 15 November 2019, viewed 25 January 2021, https://blogs.lse.ac.uk/politicsandpolicy/ge2019-pm-and-the-pendulum/

Lebo, M., & Norpoth, H. (2007), 'The PM and the Pendulum: Dynamic Forecasting of British Elections', British Journal of Political Science 37(1): 71–87.

Lewis-Beck, M. S. (1985), 'Election Forecasts in 1984: How Accurate Were They?', PS: Political Science and Politics 18(1), 53–62.

Lewis-Beck, M. S. (2005), 'Election Forecasting: Principles and Practice', British Journal of Politics and International Relations 7, 145–164.

Lewis-Beck, M. S., Nadeau, R., & Bélanger, É. (2016), 'The British general election: Synthetic forecasts', Electoral Studies 41, 264-268.

Lewis-Beck, M. S., & Tien, C. (2016), 'Election Forecasting: The Long View'. In: Oxford HandBooks Online, Oxford: Oxford University Press, DOI: 10.1093/oxfordhb/9780199935307.013.92

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017), Statistical Intervals: A Guide for Practitioners and Researchers, second edn, Hoboken, New Jersey: John Wiley & Sons, Inc.

Murphy, A. H. (1992), 'Climatology, persistence, and their linear combination as standards of reference in skill scores', Weather and Forecasting 7(4), 692–698.

Murr, A. E. (2015), 'The Party Leadership Model: An early forecast of the 2015 British General Election', Research & Politics 2(2), 1–9.

Murr, A. E. (2016), 'The wisdom of crowds: What do citizens forecast for the 2015 British General Election?', Electoral Studies 41: 283–288.

Murr, A. E., Stegmaier, M., & Lewis-Beck, M. (2019), 'Citizen forecasting 2019: a big win for the Conservatives', published 4 December 2019, viewed 25 January 2021, https://blogs.lse.ac.uk/politicsandpolicy/citizen-forecasting-2019-a-big-win-for-the-conservatives/

Murr, A. E., Stegmaier, M., & Lewis-Beck, M. (2021), 'Vote expectations versus vote intentions: Rival forecasting strategies', British Journal of Political Science 51(1): 60–67.

Norpoth, H. (2001), 'Primary Colors: A Mixed Blessing for Al Gore', PS: Political Science and Politics 34(1), 45–48.

Norpoth, H. (2019), 'primarymodel16', published 9 November 2019, viewed 25 January 2021, https://twitter.com/primarymodel16/status/1193181646490624000?s=20

Prosser, C. (2016). 'Do local elections predict the outcome of the next general election? Forecasting British general elections from local election national vote share estimates', Electoral Studies 44(1), 274–278.

Prosser, C., & Fisher, S. (2019). 'The 2018 and 2019 local election results suggested the Conservatives might struggle to get a majority at the next general election', published 11 December 2019, viewed 25 January 2021, https://electionsetc.com/2019/12/11/the-2018-and-2019-local-election-results-suggested-the-conservatives-might-struggle-to-get-a-majority-at-the-next-general-election/

Quinn, T. (2012), Electing and Ejecting Party Leaders in Britain, Basingstoke: Palgrave Macmillan.

Sjöblom, G. (1968), Party Strategies in a Multiparty System, Lund: Student-litteratur.

Stark, L. P. (1996), Choosing a Leader: Party Leadership Contests in Britain from Macmillan to Blair, London: Macmillan.

Stegmaier, M., & Norpoth, H. (2017), 'Election Forecasting', in: Oxford Bibliographies, DOI: 10.1093/OBO/9780199756223-0023

Stewart, M. C. & Clarke, H. D. (1992), 'The (Un)Importance of Party Leaders: Leader Images and Party Choice in the 1987 British Election', Journal of Politics 54, 447–70.

Whiteley, P., Sanders, D., Stewart, M., & Clarke, H. (2011), 'Aggregate level forecasting of the 2010 general election in Britain: The Seats–Votes model', Electoral Studies 30(2): 278–283.