# PollBasePro: Daily Estimates of Aggregate Voting Intention in Britain from 1955 to the Present Day[*]

Jack Bailey[†], Mark Pack[‡] & Luke Mansillo[§]

This version compiled 11 March 2021 using PollBasePro v0.1.0

## Abstract

Political scientists often use public opinion polls to test their theories. Yet these data present some difficulties. First, they are noisy. Second, they occur at irregular intervals. Third, they measure both public preferences and pollsters' survey design choices. We introduce a new dataset, PollBasePro, that accounts for these issues. It includes voting intention estimates for each of Britain's three largest parties on each day between the 1955 general election and the present. Given that the dataset covers 24,032 days in total, it is more comprehensive than any existing time series of public opinion in Britain. As a result, we expect PollBasePro to be an important resource for students of British politics in the years to come.

**Keywords:** Opinion polls; public opinion; British politics; Bayesian methods.

---

[†]Research Associate, Department of Politics, University of Manchester, UK. If you have any comments or questions, feel free to contact me either by email (jack.bailey@manchester.ac.uk) or on Twitter (@PoliSciJack).

[‡]President, The Liberal Democrats, UK

[§]PhD Candidate, Department of Government & International Relations, University of Sydney, Australia

## Introduction

Beleaguered politicians often remark that the "only poll that matters is the election itself." Like much that politicians say, this is not true. Though elections decide who governs, public opinion polls are important too: they let us track how popular parties are in the time between elections. Polling is especially important in Britain. Unlike in many countries, British governments have some control over scheduling elections. As a result, the governing party's standing is often one of the only factors that decides when it will hold an election (Smith 2003).

Since the advent of universal suffrage in 1928, Britain has held only 24 general elections. This is not a large sample. Political scientists have, thus, turned to public opinion polls to test their theories. Still, these data present some difficulties. Three issues are most serious. First, that polling data are noisy and may be prone to bias. Second, that they occur at irregular time intervals. And, third, that they measure both real changes in the electorate's preferences *and* the design choices of the companies that run them.

In this paper, we introduce a new dataset—*PollBasePro*—that overcomes these issues. It includes 24,032 daily estimates of aggregate voting intention for each of Britain's three largest parties on each day between 26 May 1955 and 11 March 2021. Covering 24,032 days in total, PollBasePro permits a degree of specificity and flexibility beyond that of any existing time series data. In the sections that follow, we elaborate on the dataset. First, we describe the methods and the data that we use to derive our estimates. Next, we consider what PollBasePro reveals about British politics since 1955 and show how to use it to test hypotheses. We then provide some initial conclusions and remark on how we might develop PollBasePro in the future.

## Data, Estimation, and Validation

Polling data have long informed political science research. For example, Goodhart and Bhansali (1970) used Gallup and NOP polls to show that changes in the economy affect support for the incumbent party, thereby launching economic voting research. Though plentiful, these data come with some problems. As we mention above, three are most important.

First, polling data are noisy. This is an unavoidable consequence of sampling methodologies. As we cannot poll every person in a country, polls cannot provide exact estimates of public opinion. Rather, they represent draws from a distribution of *possible* estimates of public opinion. This noise is a type of measurement error, and failing to account for it can have important consequences. Where we use polls as an outcome, it reduces our statistical power. Where we use them as a predictor, it pulls real and existing effects towards zero.

Second, polling data occur at irregular intervals. What's more, they can cover any period of time from a single day to several weeks. Political scientists often assume that these figures measure public opinion on the poll's last day in the field. Of course, this is almost never true. It can also be a problem when major events occur partway through the data collection process. This is a major practical problem: polls can be difficult to match to other time series.

Third, polls do not only measure changes in aggregate preferences. They also include systematic biases due to the design choices of the companies that run them. In the past, these biases have been so large that they have cast doubt on the efficacy of the polling industry. At the 2015 UK general election, most polls suggested that Labour had a good chance of becoming the largest party. But, on the night, the Conservatives won a small majority instead Mellon and Prosser (2017). These biases were even more serious at the 2019 Australian Federal Election. As in the British case, Labor had a healthy lead in the polls. Yet, again, the Coalition beat them by around 8 percentage points (Mansillo and Jackman 2020).

Our intention is simple: to produce voting intention estimates that account for these issues. To do so, we adapt the method in Jackman (2005). Jackman's model has estimates for a given party start and end at known results from any given pair of elections. It then treats the party's level of support as a random walk between these two points. On any given day, the party's support depends on its support the day before, pollster-specific "house" effects, and random shocks. Jackman and Mansillo (2020; 2016) have used the model to track voting intention in Australia and Louwerse (2016) have used it to estimate aggregate voting intention in Ireland. The full model specification is as follows[1]:

---

[1]Note that we provide a more in-depth explanation of how the model works and the steps that we have taken to produce our estimates in the user guide that accompanies the data.

$$Poll_i \sim \text{Normal}(\mu_i, \sqrt{\sigma^2 + S_i^2}) \qquad \text{Likelihood function}$$

$$\mu_i = \alpha_{Day[i]} + \delta_{Pollster[i]} \qquad \text{Measurement model on } \mu$$

$$\alpha_t = \alpha_{t-1} + \tau\omega_{t-1} \text{ for } t \text{ in } 2, ..., T-1 \qquad \text{Dynamic model on } \alpha_t$$

$$\alpha_T \sim \text{Normal}(\alpha_{T-1}, \tau) \qquad \text{Adaptive prior on } \alpha_T$$

$$\delta_j \sim \text{Normal}(0, 0.1) \text{ for } j \text{ in } 1, ..., J \qquad \text{Prior on house effects, } \delta$$

$$\omega_t \sim \text{Normal}(0, 0.1) \text{ for } t \text{ in } 1, ..., T-1 \qquad \text{Prior on random shocks, } \omega$$

$$\tau \sim \text{Normal}(0, 0.05)^+ \qquad \text{Positive prior on scale of innovations, } \tau$$

$$\sigma \sim \text{Exponential}(20) \qquad \text{Prior on residual error, } \sigma$$

Two data sources inform our estimates. The first is the PollBase dataset of historic British voting intention polls (Pack 2021). The second is data compiled by volunteers on Wikipedia (for an example of this data, see Wikipedia 2021). Both are comprehensive, high-quality, and track British public opinion over the past several decades.

The PollBase data come from a range of sources. This includes books published after each general election, polling almanacs, data shared amongst academics and pollsters, contemporary media reports, and figures from polling company websites. The data since the 1983 general election are complete, baring any individual errors. Likewise, data are complete for every general election campaign before 1979. The coverage between elections is more comprehensive for some pollsters than others[2]. Gallup and NOP data, in particular, are well-covered as the two companies have collated and published their results. The PollBase data start on 24 June 1945. Yet, consistent polling began only after the 1955 general election. Thus, we take it as our starting point.

We use the Wikipedia data to cover the period from 2010 onwards. This is a pragmatic choice. Volunteers update the data in real time and include sample sizes that are missing from PollBase.

---

[2]We encourage any readers who might know of polls missing from the PollBase dataset to contact its author via his website (https://www.markpack.org.uk)
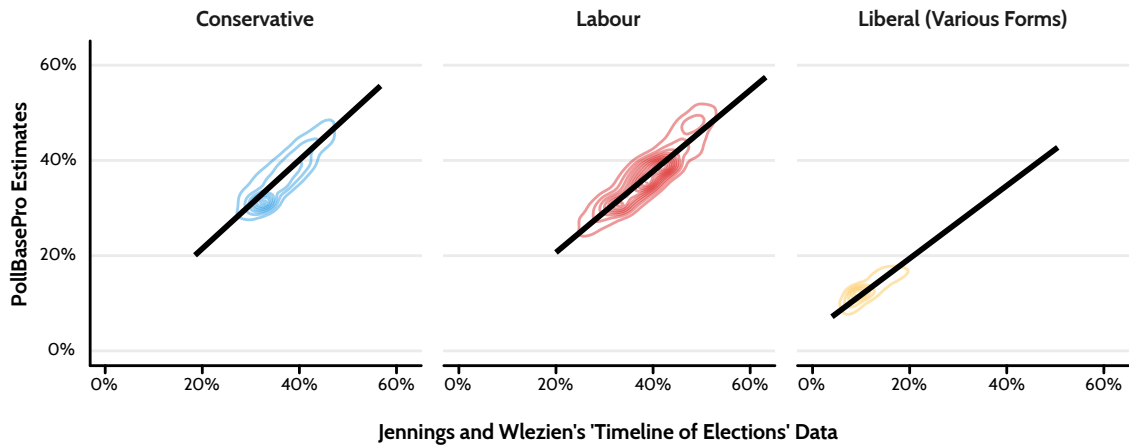
Figure 1: In all cases, estimates from PollBasePro appear well-validated when compared to raw polling data from Jennings and Wlezien's 'Timeline of Elections' dataset (2016).

This is important as our model assumes that we know the sampling error present in each estimate. For the PollBase data, we impute likely sample sizes to allow us to estimate these errors[3]. But as the Wikipedia data include sample sizes, imputation is not necessary. While some might doubt Wikipedia's reliability, we do not think that it is an issue. Polling figures are verifiable and likely of interest only to a very small group of people. Wikipedia's coverage of public opinion polling since 2010 is also of a very high quality, and almost all figures on the website including links to external data sources that corroborate them.

We validate them against Jennings and Wlezien's "Timeline of Elections" dataset (2016). These data contain 4,302 polls from Britain from 15 June 1943 to 6 June 2017. Given that our data are so comprehensive, it is likely that most polls appear in both datasets. Still, the Timeline data provide a good test as Jennings and Wlezien compiled them independently. As figure 1 shows, our estimates are well validated. Correlations between the two series are strong and positive. Their mean absolute error (MAE) and root-mean-square error (RMSE) are also low in all cases. The Conservatives showed a correlation of 90.1% (95% CI: 89.5% to 90.7%), an MAE of 2.18 percentage points, and an RMSE of 0.03; Labour, a correlation of 92.0% (95% CI: 91.5% to 92.4%), an MAE of 2.89 points, and an RMSE of 0.04; and the Liberals, a correlation of 91.6% (95% CI: 91.1% to 92.1%), an MAE of 2.22 points, and an RMSE of 0.03.

---

[3]Again, see the user guide that accompanies the data for more information on this process.

## PollBasePro and British Politics Since 1955

Figure 2 shows that our estimates track British political history very well. Between 1955 and 1980, we observe the heyday of the two-party system. Here, around 42.4% of the electorate supported Labour and another 44.8% the Conservatives. In the 1980s, we see the rise and fall of the SDP-Liberal alliance. Labour's slow rise to power in 1997 soon follows, as does their loss of support over the next decade and a half. More recently, the data show a brief "blip" in Liberal support to coincide with "Cleggmania" in 2010, Labour's surprise surge in 2017, and the Conservative's landslide support in 2019.

Table 1 summarises our estimates. One fact jumps out: that Labour and the Conservatives were almost perfectly matched. Each averaged support from around 40% of voters, this support varied by around 7 percentage points, and each took the lead around 50% of the time. The Liberals—Britain's third most popular party for much of the period—were not so fortunate. Their support averaged around 14%, though this figure varied between a low of 2% and a high of 47%, with the party taking the lead only around 1% of the time.

As well as summarising facts about the parties, our estimates also allow us to make general claims about the relationships that exist *between* the parties. As figure 3 shows, these are all negative. Consider the effect of rising Liberal support on the Conservatives and Labour. The correlation between the Liberals and the Conservatives is -39.9% (95% CI: -40.9% to -38.8%) and between the Liberals and Labour -33.2% (95% CI: -34.3% to -32.1%). Contrary to popular arguments that the Liberals serve to split the centre-left vote (e.g. Jenkins 2019), our figures suggest that rising Liberal support has tended to hurt the Conservatives more than Labour. Interestingly, the correlation between Labour and the Conservatives is also the smallest, at -29.2% (95% CI: -30.4% to -28.0%). This is perhaps unsurprising, given that the two parties represent different sides of the left-right spectrum. Even so, that both parties have tended to target the other's vote still makes strategic sense: each party's pool of potential voters is almost always larger than the pool of potential Liberal voters.
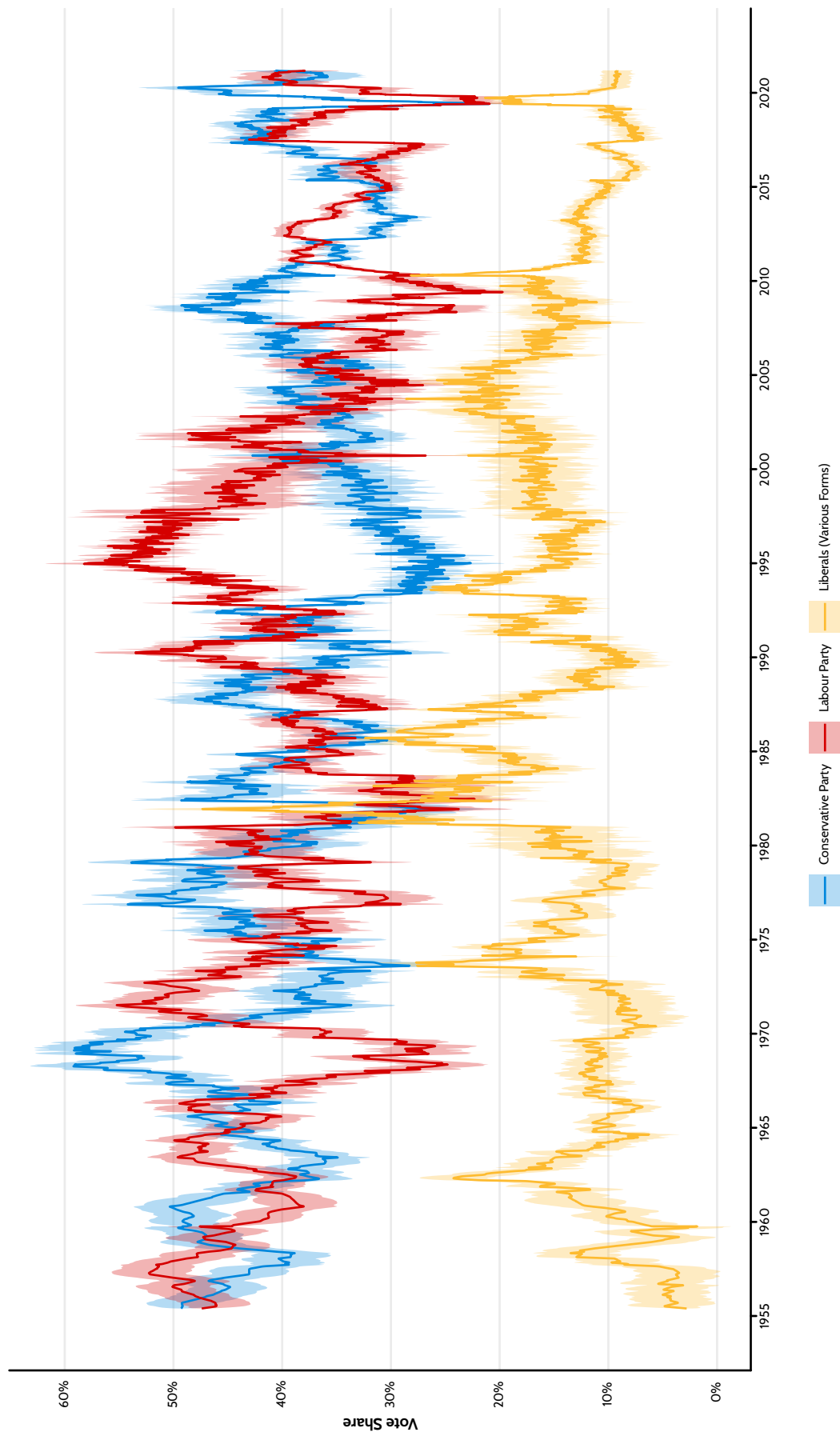
Figure 2: PollBasePro includes 24,032 daily estimates of aggregate voting intention for each of Britain's largest parties: the Conservatives, Labour, and the Liberals in their various forms.

Table 1: Overall summary of voting intention estimates, 1955 to 2021

| Party | Median | Mean | Error | Lowest | Highest | Leader |
|---|---|---|---|---|---|---|
| Conservative | 39.2% | 39.5% | 6.8% | 21.0% | 59.2% | 49.7% of days |
| Labour | 39.4% | 39.5% | 7.1% | 19.7% | 58.2% | 49.5% of days |
| Liberal | 13.5% | 14.4% | 5.8% | 1.8% | 47.3% | 0.8% of days |

We can also use our estimates to make specific claims about British politics. For example, we can assert with a reasonable degree of certainty that the Conservative Party received the largest degree of support of any party in Britain between 1955 and 2021 on 8 April 1968 when our estimates show that 59.2% (95% CI: 55.3% to 63.1%) of the electorate would vote for them at the next election. Similarly, we can assert that Labour's peak in the polls came on 19 December 1994 when 58.2% (95% CI: 55.2% to 61.3%) of the electorate intended to back them.

These data and their summaries raise an interesting question: if Labour and the Conservatives have tended to be so well-matched in the polls, why have the Conservatives done so much better at election time? Despite leading in the polls 49.5% of the time, Labour has gained the highest share of the vote in only 6 of the period's 18 general elections (33%). This phenomenon is not entirely unprecedented. Jackman (1994) shows that the Australian Labor Party has suffered at elections due to pervasive electoral bias. Though this might also be the case in Britain, our figures reflect *vote* shares, not *seat* shares. As such, electoral bias should be a concern only insofar as it affects the parties' overall popularity. Other factors likely explain Labour's poor performance: Britain's press leans right and tends to support the Conservatives; successive Conservative governments could have scheduled elections to maximise their chances; partisan non-response could have led Conservative voters to drop out of polls in the period between elections; or, most simply, the Conservatives might just be better at campaigning.

## Using PollBasePro to Answer Political Questions

It is one thing to show that our estimates are well-validated; it is quite another to show that they can advance our understanding of politics. As we mention above, raw polling data cover irregular time intervals and can be difficult to match to other time series. Our data face no such problem.
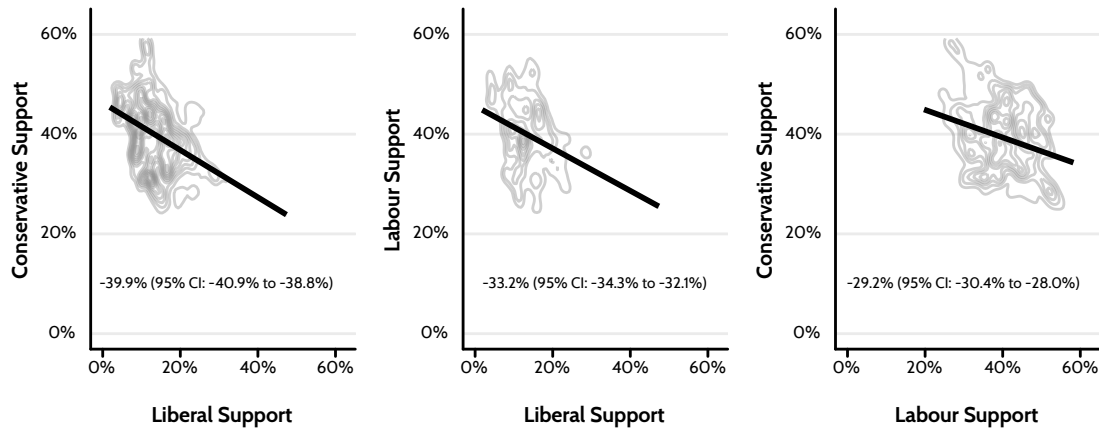
Figure 3: Historic correlations between each of Britain's three largest parties, 1955 to 2021

Instead, as PollBasePro includes daily voting intention estimates, it is simple to merge the data into any other time series, whether they be daily, weekly, monthly, quarterly, or yearly.

To demonstrate their usefulness, we use our data to answer a question of timely importance: how are deaths attributable to COVID-19 related to public support for the governing Conservative Party? To answer this question, we merge our data into the UK Government's tracker of daily COVID-related deaths (2021a). These data run from 29 February 2020 to 5 March 2021, though we begin our analysis at the 2019 General Election and mark any deaths before the start of the pandemic as zero. We then compute daily changes in Conservative support and daily changes in COVID-related deaths across a four week window, before modelling the former as a function of the latter using the following error-in-variables model:

$$\Delta C_t \sim \text{Normal}(\mu_t, \sqrt{\sigma^2 + S_t^2}) \qquad \text{Likelihood function}$$

$$\mu_t = \alpha + \beta_1 T_t + \beta_2 \Delta D_t + \beta_3 T_t \times \Delta D_t \qquad \text{Linear model on } \mu$$

$$\alpha \sim \text{Normal}(0, 0.1) \qquad \text{Prior on } \alpha$$

$$\beta_j \sim \text{Normal}(0, 0.1) \text{ for } j \text{ in } 1, ..., J \qquad \text{Prior on } \beta$$

$$\sigma \sim \text{Exponential}(10) \qquad \text{Prior on } \sigma$$

This is a simple linear model, where change in Conservative support at time $t$, $\Delta C_t$, is a function of the passage of time, $T_t$, and change in the number of COVID-related deaths, $\Delta D_t$. We also interact the two variables so that the effect of these deaths can change over time. As the intervals in figure 2 make clear, our data are *probabilistic*. Thus, it is important that we propagate this uncertainty forward so as to maximise our statistical power. This is what differentiates our error-in-variables model from a standard linear model: rather than include only the residual error, $\sigma$, we also include any known uncertainty in our estimates at time $t$, $S_t$.

According to the literature, there are good reasons why the relationship between the two variables might be positive and good reasons why it might be negative. The literature on "rally 'round the flag" effects would favour a *positive* relationship. Here, voters rush to support the governing party in times of crisis, such as wars (Kuijpers 2019; Mueller 1970). The literature on retrospective voting (Healy and Malhotra 2013; Fiorina 1981), instead, would favour a *negative* relationship. It holds that voters reward governments for positive policy outcomes and punish them for negative ones.

Figure 4 shows that the effect is more "rally 'round the flag" than "reward and punishment." That said, our model suggests at least some preliminary evidence of a transition from one mechanism to the other. In March 2020, when the pandemic started to gain pace in Britain, there was a strong positive relationship between change in the number of deaths attributable to COVID-19 and change in support for the governing Conservative Party. For every 100 extra COVID-related deaths over the past 4 weeks, Conservative support over the same period increased on average by 0.6 points (95% CI: 0.5 to 0.8). Six months later, in September 2020, this relationship had weakened markedly. Though still positive, it had fallen to only 0.2 points (95% CI: 0.2 to 0.3). A year after the pandemic began, the relationship had reversed. In March 2021, every 100 additional COVID-related deaths over the past 4 weeks led Conservative support to fall on average by -0.2 points (95% CI: -0.2 to -0.1).

It is clear then that PollBasePro has allowed us to reveal an important—if surprising—answer to the question that we posed above: despite Britain having one of the world's highest levels of COVID-related deaths per capita (Ritchie et al. 2021), these untimely deaths appear to have
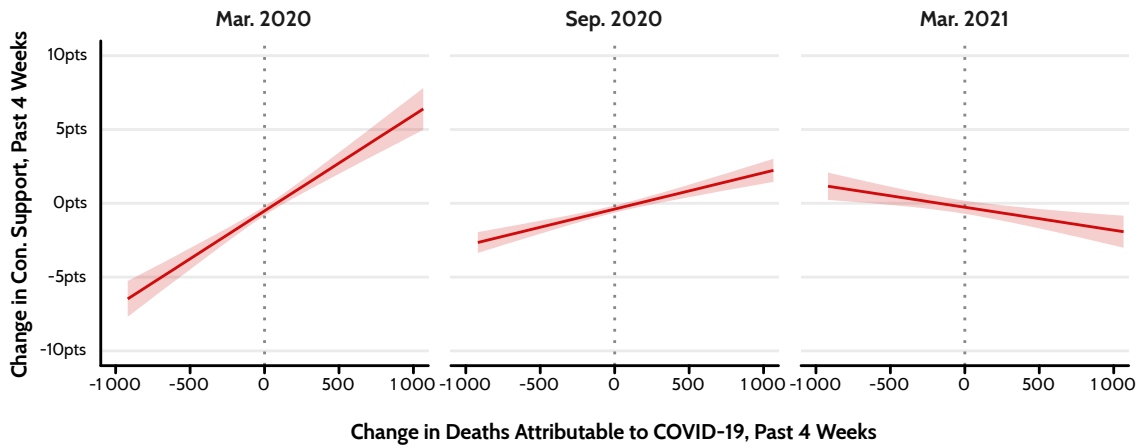
Figure 4: At the start of the COVID-19 pandemic, more deaths were related to greater support for the incumbent Conservative Party. As time has passed, this relationship has reversed.

*benefited* the fortunes of the governing Conservative Party. Still, the relationship has changed over time and, a year into the pandemic, has reversed direction. We might, thus, expect any further deaths attributable to COVID-19 to erode the Conservative's ample base of support. Yet given the pace of Britain's vaccination programme (UK Government 2021b), it is possible that the Conservatives have reaped the rewards without facing any punishment.

## Conclusions

In this paper, we introduce PollBasePro: the most comprehensive time series of British voting intention data assembled to date. The dataset contains 24,032 daily voting intention estimates for each of Britain's three largest parties, beginning at the 1955 General Election on 26 May 1955 and ending on 11 March 2021. Furthermore, our data are well validated and follow the course of British political history. We therefore expect PollBasePro to become a valuable resource for students of British politics in the years to come.

It is important to stress that PollBasePro is a *living dataset*. As a result, we expect to incorporate new estimates into the data as time goes by. Users should, therefore, endeavour to use the most recent data in any of their analyses. It is also important to stress that—like any project of this nature—our data pipeline might contain minor errors or inefficiencies. To guard against this, we have hosted all of our materials online for others to inspect. If our users find any

errors in our code or wish to make recommendations for future updates, we invite them to raise an issue on the project's GitHub repository or to contact the authors directly.

Though we believe that our data can help to answer all manner of questions, PollBasePro still has room to grow. Three new features would be most useful. First, to estimate aggregate support for the UK leaving the European Union and for Scotland leaving the UK. Both time series would be long-ranging and benefit from having known outcomes to which we could anchor our estimates. Second, to expand the data beyond the three main parties. This would be particularly useful for those who study, for example, the rise of the SNP in Scotland or the influence of UKIP. Third, to collect and incorporate known sample sizes into all of our polling data. At the moment, we use the Timeline of Elections dataset (Jennings and Wlezien 2016) to impute sample sizes for all polls that occurred before the 2010 general election. This is a pragmatic but reasonable decision given that these data do not include sample sizes. Still, were we to have the necessary resources, it would be good to collect the sample sizes associated with these polls to ensure that our estimates are as precise as possible.

# References

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Goodhart, C. A. E., and R. J. Bhansali. 1970. "Political Economy." *Political Studies* 18 (1): 43–106. https://doi.org/10.1111/j.1467-9248.1970.tb00659.x.

Healy, Andrew, and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16 (1): 285–306. https://doi.org/10.1146/annurev-polisci-032211-212920.

Jackman, Simon. 1994. "Measuring Electoral Bias: Australia, 1949." *British Journal of Political Science* 24 (3): 319–57. https://doi.org/10.1017/S0007123400006888.

———. 2005. "Pooling the Polls over an Election Campaign." *Australian Journal of Political Science* 40 (4): 499–517. https://doi.org/10.1080/10361140500302472.

Jackman, Simon, and Luke Mansillo. 2016. "The Campaign That Wasn't: Tracking Public Opinion over the 44th Parliament and the 2016 Election Campaign." In *Double Disillusion: The 2016 Australian Federal Election*, edited by Anika Gauja, Peter Chen, Jennifer Curtin, and Juliet Pietsch, 720. Acton, AU: ANU Press.

Jenkins, Simon. 2019. "The Lib Dems Helped the Tories to Victory Again. Now They Should Disband." *The Guardian*. https://www.theguardian.com/commentisfree/2019/dec/16/lib-dems-tories-split-vote-labour.

Jennings, Will, and Christopher Wlezien. 2016. "The Timeline of Elections: A Comparative Perspective." *American Journal of Political Science* 60 (1): 219–33. https://doi.org/10.1111/ajps.12189.

Kuijpers, Dieuwertje. 2019. "Rally Around All the Flags: The Effect of Military Casualties on Incumbent Popularity in Ten Countries 1990." *Foreign Policy Analysis* 15 (3): 392–412. https://doi.org/10.1093/fpa/orz014.

Louwerse, Tom. 2016. "Improving Opinion Poll Reporting: The Irish Polling Indicator." *Irish Political Studies* 31 (4): 541–66. https://doi.org/10.1080/07907184.2016.1213719.

Mansillo, Luke, and Simon Jackman. 2020. "National Polling and Other Disasters." In *Morrison's Miracle: The 2019 Australian Federal Election*, edited by Anika Gauja, Marian Sawer, and Marian Simms, 125–48. Canberra, AU: ANU Press.

Mellon, Jonathan, and Christopher Prosser. 2017. "Missing Nonvoters and Misweighted Samples." *Public Opinion Quarterly* 81 (3): 661–87. https://doi.org/10.1093/poq/nfx015.

Mueller, John E. 1970. "Presidential Popularity from Truman to Johnson." *The American Political Science Review* 64 (1): 18–34. https://doi.org/10.2307/1955610.

Pack, Mark. 2021. "POLLBASE: OPINION POLLS DATABASE FROM 1943-TODAY." https://www.markpack.org.uk/opinion-polls/.

Prosser, Christopher, and Jonathan Mellon. 2018. "The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses." *Government and Opposition* 53 (4): 757–90. https://doi.org/10.1017/gov.2018.7.

Ritchie, Hannah, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, and Max Roser. 2021. "Coronavirus (COVID-19) Deaths." *Our World in Data*. https://ourworldindata.org/covid-deaths.

Smith, Alastair. 2003. "Election Timing in Majoritarian Parliaments." *British Journal of Political Science* 33 (3): 397–418. https://doi.org/10.1017/S0007123403000188.

Sturgis, Patrick, Jouni Kuha, Nick Baker, Mario Callegaro, Stephen Fisher, Jane Green, Will Jennings, Benjamin E. Lauderdale, and Patten Smith. 2018. "An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181 (3): 757–81. https://doi.org/10.1111/rssa.12329.

UK Government. 2021a. "Deaths Within 28 Days of Positive Test by Date of Death." *GOV.UK*. https://coronavirus.data.gov.uk/details/deaths.

———. 2021b. "People Who Have Received Vaccinations, by Report Date (daily)." *GOV.UK*. https://coronavirus.data.gov.uk/details/vaccinations.

Wikipedia. 2021. "Opinion Polling for the 2019 United Kingdom General Election." *Wikipedia*. https://w.wiki/35RK.