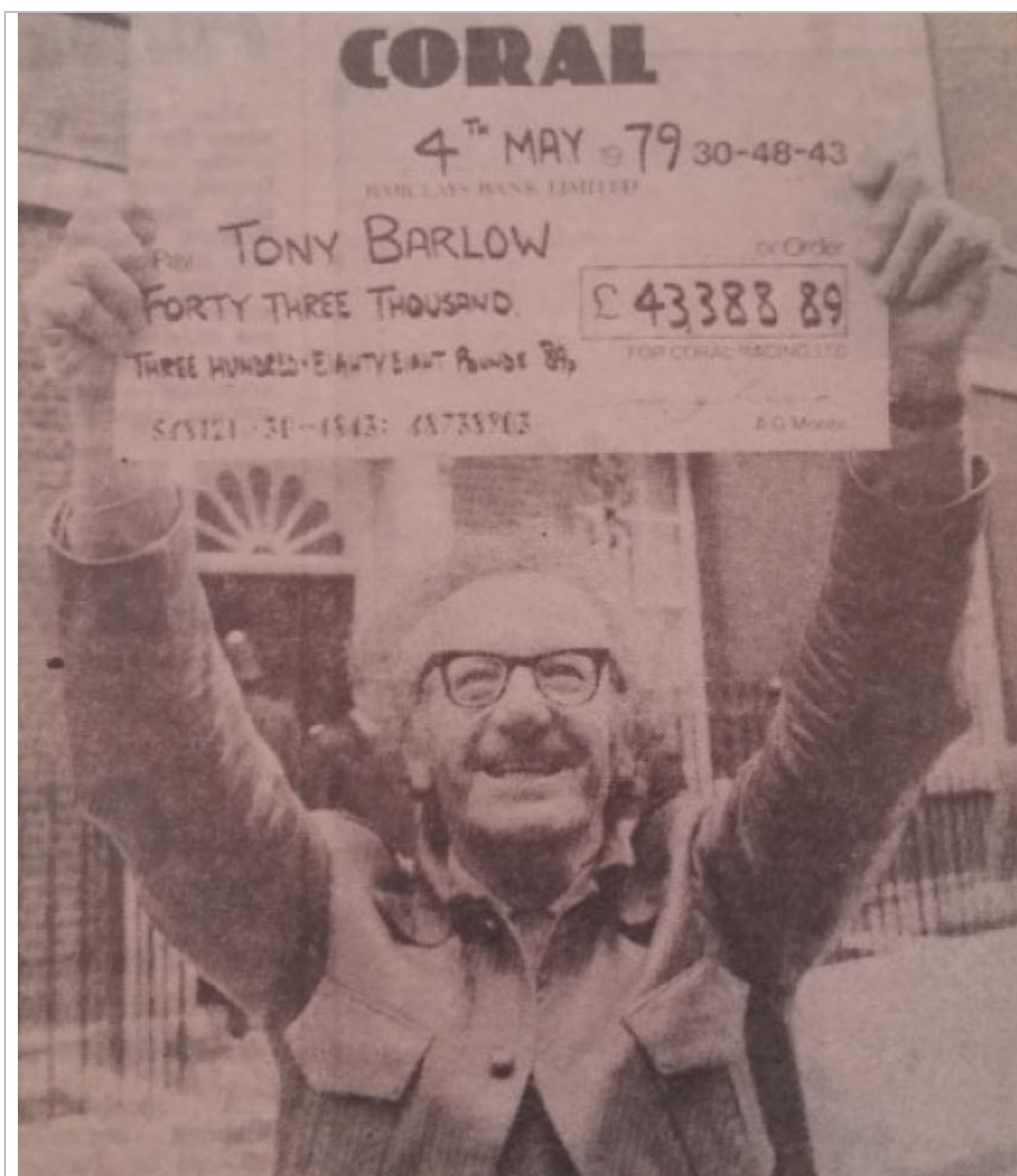# FORECAST ERROR: HOW TO PREDICT AN ELECTION, PART 2: BETTING
"Bet or be silent."[0401a]

Tony Barlow holding aloft his winnings from Coral following his £30,000 bet on the Conservatives winning the 1979 election. Photo from the Sidmouth Herald website, published 14:35 28 July 2016

1. PREAMBLE
In 1979, Mr Tony Barlow was interviewed by the BBC. He had a nice house in the suburbs, two cars, four children and a wife with a long bowl-cut. He also had a gambling habit: a big one. He'd bet £30,000 at 4/9 that the Conservative party was going to achieve a majority at the 1979 UK General Election.[0707a]

The average price of a house in Q2 1979 was £19,075[0416e]. He seemed relaxed when speaking to the interviewer, and later bet more money.

He won.[0416f]

Welcome to the world of political betting in the United Kingdom. It is a discipline with showmanship. It excites passions and disapproval. It involves misdirection and elision, great study and great neglect. It is fervently believed and yet disdained. Its centuries-long course has matched the evolution of the law, the class struggle and the country itself. It covers wars, famines, Chief Justices, and a teenager who became Prime Minister. And this article attempts to describe it.

2. INTRODUCTION

The "Forecast Error" series of articles started examining election predictors in 2015. Each article considered many predictors, but each article covered just one election. This article is the second in a new chapter in the "Forecast Error" series where we examine an individual class of predictor more closely across many elections. This article covers political betting in the United Kingdom.

As this is the UK, the word "punter" will be used to describe that entity which places the bet on the future event, and "bookmaker" or "bookie" for that entity which holds the money until the bet is resolved. Terms such as "turf accountants" will be avoided where possible. The word "gaming" occurs too often in history to ignore but our preference is for "betting" or "gambling". The "odds" is the ratio of the money staked on an outcome to the profit if the outcome happens, and it can be rendered as a probabilistic prediction. The "book" or "full book" will be used to describe the set of all odds offered on a given election at a given moment. The "overround" is the slight change in the odds imposed by the bookie to ensure a profit, and has to be removed to stop the book probability exceeding one. "Short odds" (eg 1/100) is used to describe events thought to have a high probability, "long odds" for those with a low probability.

3. SCOPE

Which brings us to an awkward question: fine, but how do we define "betting" or "gaming" or "gambling"? This question is superficially easy, but not in real life it isn't. Consider the various discussions on the differences between "gaming", "gambling" and "betting"[0518t] or the fuzzy boundaries between the stock market, futures trading, prediction markets and gambling[0518u]. They've been frequently defined in law, for example "*a promise to give money or money's worth upon the determination of an uncertain or unascertained event in a particular way*"[0416a]. But that definition would cover everything from fixed-odds betting, spread betting, betting exchanges, share dealing, futures trading, prediction markets, probability consultancy, insurance and actuarial studies. So how to resolve this? What shall we use as a sieve?

Fortunately there is a simple but brutal sieve available to us: "*it's gambling if the Government of the United Kingdom, or its constituent parts or predecessors, said so*". This sieve is available to us because of the Gambling Act 2005, which provided a definition of "gambling"[0518l] and set up the Gambling Commission. If you offer odds to somebody in Britain, you have to obtain a licence and be regulated by the Gambling Commission, regardless of where you are in the world. So, having obtained our sieve, what shall we cover?

This article will cover fixed-odds betting at national level, aka "sportsbook", aka "High Street betting" (legally 1960-date)), "Majorities" betting (1906?-1935), Club and College betting (18th century to date) and Exchange betting (21st century). It will not cover other things that are defined as gambling such as constituency betting. It will not cover things that are thought of as gambling but which are not caught by our sieve, such as prediction indices and spread betting: those are discussed in the Appendices and will be covered by later articles

4. HOW TO ASSESS ACCURACY

Here's a question: how exactly do you assess the goodness of a probabilistic prediction? In previous articles we used MAE and WIN, but is there anything else that can be used? Over the course of the "Forecast Error" the question has frequently arisen and alternative metrics have been suggested. A previous article in the series went thru several metrics for assessing the proximity of a string of values to another paired string[0602a], and conversations with colleagues yield still more. So, which to use?

For this article we interviewed bookies and academics and looked at the available literature. Interviews with Matthew Shaddick (head of Political Betting, Ladbrokes) and Graham Sharpe (former Media Relations Director, William Hill) pointed out that generally speaking, accuracy is not formally measured post facto, although gross mispricings are inspected and faults corrected if found, but Shaddick also made a passing reference to calibration. When considering the work of academics, Professors Christopher Wlezien (Hogg Professor of Government at the University of Texas at Austin) and Leighton Vaughan Williams (Nottingham Business School) gave an interview and factsheet respectively. They used distance metrics such as MAE although Vaughan Williams also used accuracy, bias and precision[0602b]. A consideration of the literature covered the work of Vaughan Williams and Reade[0602b][0601b][0602g] and Fisher[0602c] and Smith and Gustafson[0602d] and Erikson and Wlezien[0602e][0602f][0602h] and others[0602i][0602j][0602k][0602l][0602m], which yielded further metrics. Lastly we considered work done with measuring the goodness of weather forecasting[0602o], reasoning that metrics used to measure the goodness of probabilistic weather forecasts could be used for probabilistic election forecasts. After consideration, we chose the following.

To retain compatibility with previous articles we will continue to use MAE and WIN to measure accuracy (or more subtly predictiveness) of the book. "MAE" is Mean Absolute Error, the average absolute difference between the forecast probability and actual probability for each odds in a book (the formula is $MAE = (\sum |Pi - Oi|)/N$, where N is the number of odds in a book), and "WIN" is a simple binary value: 1 if the book predicted the winner, 0 if it didn't.

In deference to the work of Vaughan Williams and Shaddick we will also measure bias and calibration of the individual odds. We considered measuring precision of the book, but it was too similar to MAE to add meaning. We will also calculate the Brier Score, although its use for multiple nonbinary books raised questions regarding thresholds, which we discuss below. For full definitions of "MAE", "WIN", "bias", "calibration" and "Brier score", see "Appendix 1: Metrics used in this article to assess the goodness of probabilistic forecasts".

Please note that an interesting phenomenon became apparent whilst reading the academic papers listed above. In many cases the elements measured were not the raw data itself but were instead transformed data derived from the raw data by models, regressions or other calculations. This approach

enables the academics to draw more sophisticated conclusions from the data and they have well-argued reasons for this. But this makes the work difficult to reproduce and distances the metric from the thing we want to measure. Consequently, our measurements will be based on the raw data, with the only transformation being the removal of the overround where possible. In deference to Wlezien's terminology[0602e] we will distinguish our approach from the more sophisticated academic one by describing our approach as the *naïve reading*. A more in-depth analysis of the academic approaches will be deferred to a separate article, provisionally entitled "Forecast Error: Academics and analysts"

5. PREDICTIVENESS VERSUS ACCURACY

At this point we have to discuss the difference between predictiveness and accuracy. This was discussed in our previous article on polling[0819a] but we reprise it here. To measure predictiveness as distinct from accuracy you must define your metric without knowing the election results beforehand. But when a surprise happens (eg the Scottish SNP gaining 56 out of 59 seats in Scotland in 2015) the natural human temptation is to measure against that.

Unfortunately, this leads to comparison problems: you end up tuning your metric against the SNP and UKIP in 2015, the SDP in 1983…and suddenly you don't have one consistent metric across many elections, you have many different metrics each predicting a different election. To avoid this problem we need a robust metric that will cope with many elections consistently.

To that end we shall use CON/LAB/OTHER - "three-party-forced" - to judge accuracy for our MAEs. What we lose in detail we gain in robustness as this combination will cope with most UK General Elections since 1906. Books which offered odds on fewer than three parties, such as the pre-1945 "majorities" market, will use three-or-less parties.

6. THRESHOLD

Here's a question: how good is good enough? How bad does a book or an odds have to get before we say it's too bad? What is our threshold?

In a previous article[0819a] we discussed thresholds and noted that polls were judged using a 2% MAE threshold for four-party-forced, which equates to a 2.7% MAE threshold for three-party-forced. So we will similarly use 2.7% or less as our threshold for MAE three-party-forced for this article. For WIN, it will be simpler: did a majority of the polls predict the correct winner?

In modern days bookmakers dislike[0819b] setting very short odds and very long odds, and the days when Ladbrokes would offer 50,000 to 1 against the Liberals winning the election as it did in 1964[0709x] are long gone. This imposes a structural bias on odds: even events with an extraordinarily low/high probability will not be assigned extraordinarily long/short odds. Consequently we will not impose a threshold on bias, instead merely describe it.

We calculated two versions of the Brier Score: the "Brier Score across all books" and the "ungrouped Brier Score", both on a Lab/Con/Oth basis with the overround removed. The former groups the individual Lab, Con and Oth predictions into books, the latter considers them ungrouped. The ungrouped Brier Score uses the number of possible outcomes as the divisor: under that formulation the Brier Score equates to the Mean Squared Error and hence varies from 0 (good) to 1 (bad). Most

examples of the Brier Score we found online refer to that formulation. The "Brier Score across all books" is the original Brier Score as defined by Brier in his 1950 paper[0602p], and that uses the number of books as the divisor, not the number of possible outcomes. Using that method the maximum score may exceed 1, as Brier demonstrates[0602p] in his paper (*From consideration of this example or of formula (2), it is obvious that the score P has a… maximum value of 2 for the worst possible forecasting*").

Brier Scores are used to compare predictions and an absolute threshold is not known to us. Consequently we will not impose a threshold on the Brier Score, instead merely describe it.

7. GEOGRAPHY

The United Kingdom currently holds the longform name of "The United Kingdom of Great Britain and Northern Ireland"[0416c] and is presented as a country of four countries: England, Scotland, Wales and Northern Ireland. But its borders and constituents have changed drastically and repeatedly in the past and that matters when you are dealing with legislation in the times of King Henry VIII and case law at the time of King George III. Writings on gambling frequently elide "the UK" to "Great Britain" and thence to "England" which can lead to problems.

Sometimes these elisions just lead to misunderstanding: for example the Gambling Act of 2005 is only partly applied to Northern Ireland, where the Betting, Gaming, Lotteries & Amusements (NI) Order 1985 still mostly holds sway and bookmakers do not open on Sunday. Sometimes the elisions have a tragic dimension that require historical context: discussion of Henry VIII of England's Unlawful Games Act of 1541 is improved by mentioning that he would invade Scotland the following year, and the gap in political betting newspaper coverage in Ireland[0401a] between betting on Parliament before 1830, and betting on US/Canadian elections at the end of the nineteenth century, needs to take into account the nineteenth-century famines in Ireland and the resultant diaspora. This article will attempt to deal with this long history by mentioning the kingdoms and Parliaments involved at the time.

So, let us now look at that history. A longer version can be found in the article "Forecast Error: Supplement 1: The History Of Political Betting", but the short version is given below.

8. HISTORY

1190-1707: KINGS, PRINCES AND BARONS

Let's start with the Medieval kings. King Richard I of England and King Philip II of France prohibited (1190) "any person in the army, beneath the degree of knight, from playing at any sort of game for money : knights and clergymen might play for money, but none of them were permitted to lose more than twenty shillings in one whole day and night, under a penalty…to be paid to the archbishops…" Later decrees by King Richard II of England (1388) and King Edward IV of England (1477) directed people to practice archery and put aside gaming. In 1541 King Henry VIII of England decreed the Unlawful Games Act across the Kingdom of England and the Kingdom of Ireland, and later tried to invade Scotland. He failed, but in 1621 James VI of Scotland restricted gambling and any winnings over a certain sum would be forfeit to the kirk.

The 17th century Puritans inhibited gambling and overthrew the monarchy, but in 1660 the monarchy was restored. The Restoration King Charles II of England, Scotland and Ireland, had a passion for gambling and so gambling returned. Puritan legislation was overturned but old habits die hard and some

elements were later restored, particularly the voiding of gambling odds. Specifically the Gaming Act 1664 was introduced, making gambling debts of £100 unenforceable.

This period laid down the principles of gambling legislation: gambling was not illegal within limits between individuals of rank, but any expansion beyond that would be discouraged. Religion and class would influence legislation and dissolute scions would be prevented from gambling away their inheritance by voiding gambling debts above a certain amount. Political betting was between individuals, informal and unrecorded.

1707-1865: GENTLEMEN OF MEANS AND RANK

Queen Anne of Ireland and of the 1707 unified Great Britain had a passion for gambling and during the 18th century betting and political betting were habits of gentlemen of means in Great Britain and in Ireland, although gambling was further curtailed for the lower classes. Gentleman's clubs, colleges & coffeehouses took bets and newspapers reported on them. Such political bets included elections, battles and the lives of politicians and revolutionaries.

But this increasing prurience in the lives of others sat uneasily with Chief Justice Lord Mansfield, so in 1778 he and three others voided a bet on the basis that wagers could be voided if contrary to statute, public policy or morality. Some years later Lord Mansfield sat on "Allen v Hearn" (1785) and voided a bet on the outcome of a parliamentary constituency election.

Under the Prince Regent (later King George IV of Great Britain and Ireland) the first two decades of the 19th century saw an expansion of gambling. But betting remained a class issue: the upper classes bet in the gentlemen's clubs, but the working class bet in "copper hells" as an escape. There was a backlash, with the middle class complaining that gambling ruined families and allowed for overt class mixing. The British government enacted the Gaming Act of 1845 and other acts that increasingly limited gambling, especially among the working classes.

This period saw gambling become a matter of historical record, as bets were now written down in books that still survive. But political betting was still personal based on individuals, not party based and difficult to summarise statistically. By the end of this period the themes that had underpinned gambling for centuries had reached their apotheosis: gambling debts were now entirely void and the class distinction enforced. The copper hells would cease but the upper-class gold and silver hells would not.

1865-1935: MIDDLE-CLASS STOCKBROKERS

But it didn't turn out like that. In reality gambling in Great Britain continue to grow in the 1880's. Bookmakers simply decamped to Scotland (until 1875), France and Holland, or advertised via the press, or simply moved to the street. Towns and villages had bookmakers who set the odds, stumped up the cash, and ran a network of agents and runners. Another backlash ensued and by the 1890s the National Anti-Gambling League was formed. Acts were enacted and the numbers of arrests increased but the numbers betting remained the same.

Meanwhile on the other side of the Irish Sea in Ireland, the population had fallen considerably after the famines and the creation of a diaspora, and reports of political betting after 1890 focused on elections in the United States and Canada. In the 1920's Ireland split: the Irish Free State (later the Republic of Ireland) seceded from the UK, Northern Ireland remained within it.

By the early 20th century the London Stock Exchange was trading in "Majorities", instruments based on the size of the Government majority after the election. They were tradable and acted as a futures market. But the market collapsed after the 1931 election because the majority of 493 was so large some lost considerable sums. One broker sued, but the judge deemed Majorities contracts were void under the Gaming Act 1845. The LSE modified Stock Exchange Rule 93 to read "A member shall not deal in prospective dividends or 'Majorities'"

This period saw gambling move from the upper-classes of rank and landowners to the middle-class professionals: educated men who used their education to try to beat the odds and eventually failed. But mass gambling by the uneducated was still deprecated. Political betting had moved from betting on people to betting on parties, and meaningful statistics for a given election could finally be compiled.

1945-1974: WORKING-CLASS GAMBLERS

During World War II betting retreated to the private sphere, and public election betting was little reported, with few quotes appeared in the newspapers for the 1945 and 1950 UK General Elections. The Economist stated that "it is curious that in a nation devoted to gambling as the British, so little opportunity should nowadays be taken of a general election, the most sporting of all events". But this was naïve: the public was still gambling, just not publicly. People with credit could bet via telephone and postal betting, and those without used illegal shops via a network of runners. But things were about to change.

Royal Commissions were set up in Great Britain and Northern Ireland and betting shops were legalized in 1958 (NI) and 1960 (GB) with licenses granted in 1959/1961. Betting on political events now became trivially easy.

Betting odds were offered by Ladbrokes on the 1963 Labour party and Conservative party leadership elections, and the later general elections and presidential elections in UK, USA, France, and West Germany. Ladbrokes was joined in 1965 by William Hill after overcoming the objections of its founder, then followed by other bookmakers in later years as the market continued to expand.

This period saw gambling finally become a mass pastime available to the working population. Political betting became an acknowledged phenomenon and frequent publication of odds meant that statistics tracking movement over time could now be compiled. Politics became unstable and the outsider won in both 1970 and Feb 1974.

1974-1992: STABILISATION AND PROFIT

In 1972 the Parliament of Northern Ireland was suspended and civil administration passed to Whitehall with brief exceptions. The politics of the times were troubled, with the 1970's seeing four UK Prime Ministers, four General Elections, and four referenda on Scotland, Wales, Northern Ireland and Europe. It was a fertile time for political betting.

Politicians and bookmakers started to work political betting for publicity. In 1973 the Liberal Clement Freud stood for election. He placed several bets and brought the odds down, which generated publicity that helped him get elected. Freud then bet on the 1979 General Election and obtained a six-figure return. The bookmaker William Hill generated publicity by teaming up with Screaming Lord Such who founded the Official Monster Raving Loony Party in 1983.

In the 60's betting shops were forced to adopt a dour presentation, with odds on the walls and an audio commentary by the Exchange Telegraph Company (Extel) But in the 1980's regulation was relaxed and shops could now have seats, television and refreshments. The regulation applied to Great Britain but not Northern Ireland, which was covered under the Betting, Gaming, Lotteries & Amusements (NI) Order 1985. This will become important later.

Political betting spread from high-street bookmakers to other modes. By 1987 the spread betting firm IG Index were offering spreads, as did Sporting Index in 1992.

This period saw gambling become multimodal and easier. Comparisons of predictiveness between modes could be made. The scope of political betting widened from the limited majorities and winner options of earlier years. As politics stabilized, betting odds became good election predictors.

1997-2014: INNOVATION AND EXPANSION

Following the 1997 Labour landslide and the 1998 Good Friday Agreement, referendums were held in Northern Ireland, Scotland and Wales. After the dust had settled, some powers had been devolved to Wales, Scotland and Northern Ireland and new Assemblies and Parliaments set up.

New forms of online gambling were emerging with the rise of the online sportsbook such as sporting-life.com (later BetOnline.co.uk) and bet365.com Online exchanges also began to appear. In 2000 Betfair Exchange was launched, competing with Global Betting Exchange (later BETDAQ) and Flutter, but went on to dominate the exchange industry with competitors failing or bought out

By the late 1990's bookmakers such as Victor Chandler (later BetVictor) started to move to low-tax jurisdictions like Malta and Gibraltar. This threatened betting duty avoidance so the 2001 Budget changed bookmaker taxation to net revenue.

Existing legislation wasn't coping. The Budd Report recommended legislation be codified into a single Act, a Gaming Commission set up to regulate and license all gambling (except for spread betting), and categories created to handle internet betting. Surprisingly, it recommended that gambling debts should be legally enforceable and not void, reversing centuries of legal trends. The Gambling Act 2005 extended to England, Scotland and Wales, with some (but not all) sections extended to Northern Ireland.

This period saw gambling moving online and offshore, with legislation struggling to catch up. The enforceability of gambling debts meant that gambling could become professional. Data was widespread but proprietary: available, but at a price. Use of political betting statistics was limited to academics with institutional or private databases. Odds remained good election predictors

2014-2019: UPHEAVAL AND UNCERTAINTY

The Gambling Act 2005 had attempted to cope with betting duty avoidance but was unsuccessful, with betting companies based outside the UK working around the 2005 Act. So the Gambling (Licensing and Advertising) Act 2014 established that all operators selling into the British market, regardless of location, would need a licence. But the 1985 Order still pertained in Northern Ireland and 2013 proposals from the Northern Ireland Executive were not enacted.

BETDAQ was bought by Ladbrokes in 2013 and Betfair merged with PaddyPower in 2016. The betting exchange model had worked well but Betfair were losing money on unmatched wagers, so it opened a sportsbook in 2012 to compensate. Ladbrokes and Coral merged in 2016 and acquired by GVC in 2018

In 2016 the Gambling Commission issued a discussion paper on virtual currencies, eSports & social gaming. As of 2019, social gaming (gambling for a prize unconvertible to money or money's worth) remains unregulated & spread betting & binary options in financial markets are regulated by the FCA

This period saw gambling become data-rich. Online resources became available, either free via aggregators and online archives or cheaply via payments to bookmakers monetizing their archives. Analysis of political betting statistics became possible by the public and online discussion flourished. More unstable politics made betting odds less predictive, with multiple election favourite failures.

9. UK ACCURACY BY ELECTION

So as we saw above, we can compile statistics for the 1906-1931 Majorities market and the 1964-2017 elections. We converted the MAE's for each election to 3pf (Lab/Con/Other) without the overround, and plotted the results on a graph. We plot the following concepts:

- Average MAE: how accurate is each book on a 3PF basis?
  Polls have to be accurate to around 2.7% to be regarded a success, so how accurate are odds?
- Average WIN. Do the books predict the winner or not?
  For a given election, how many of the bookies predicted the winner
- Bias(Labour), Bias(C), Bias(Oth). How biased are the books?
  Are the odds biased in favour of Labour, Conservative or other?
- Bias(Victor). How much do the odds underestimate the chance of the eventual winner winning?
  Note that obviously they cannot overestimate them.

The graphs are below. First, we consider Mean Absolute Error: MAE. How close were the odds?

Mean Absolute Error of the book

Figure 1 (A1): Average MAE, elections 1910-1931, Con/Lab/Other basis
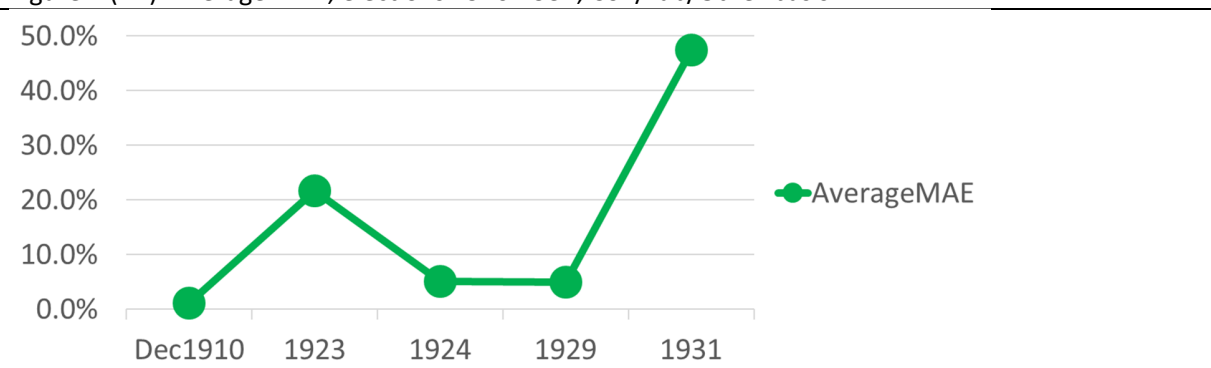
This graph depicts the average MAE of the latest available predicted majority (Dec1910, 1923, 1931) or seats (1924, 1929) compared on a three-party-forced basis to the result. The threshold is 2.7%

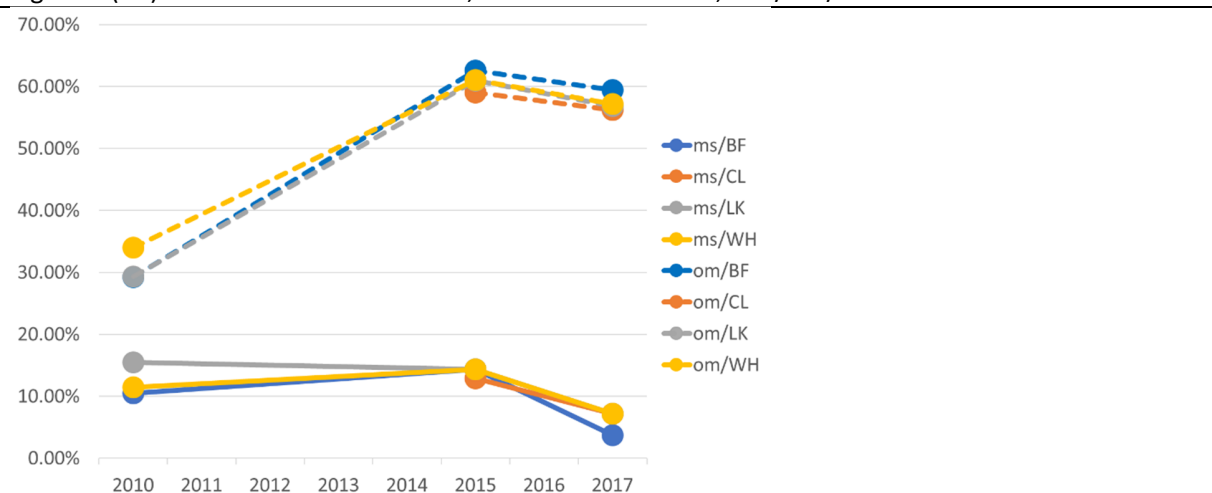Figure 2 (A2): Average MAE, elections 1964-2017, Con/Lab/Other basis
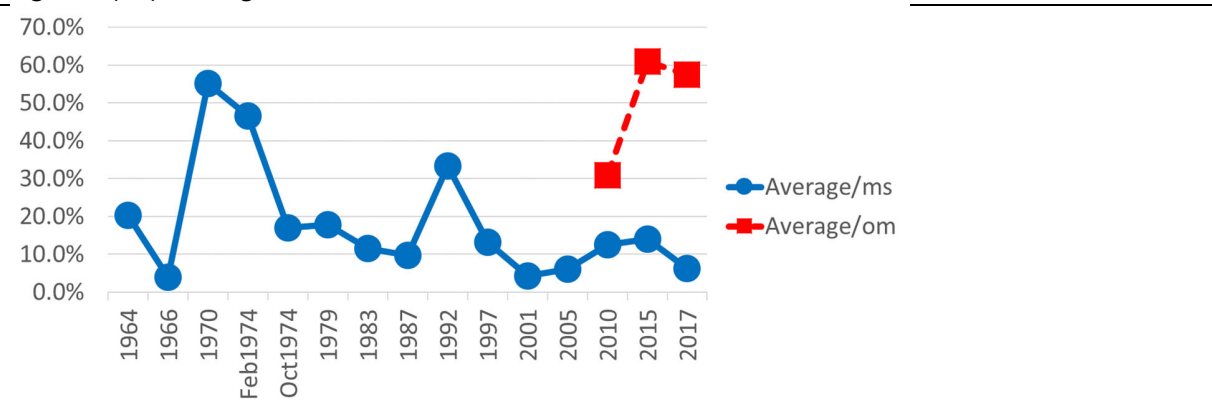
This graph depicts the average MAE of the latest available odds on most seats (solid line) or overall majority (dashed line) offered by bookies on the 1964-2017 elections compared on a three-party-forced basis to the result. The threshold is 2.7%

In our previous article on polling, we selected 2.7% as the acceptable error for multiparty predictions. Only in one previous election that we considered did it have an error under this threshold: December 1910: predicted majority of incumbent Government 130, actual 122 seats, an error of eight seats, which is 0.01% of the 1910 House of 670 seats. The rest of the time the error was above this, in some cases considerably so.

Mean Absolute Error of the book: Fixed-Odds Vs Exchange Betting

We have already investigated MAEs over elections. But how about MAEs between modes? Are fixed-odds better than exchange betting? We depict the MAE of the book for each bookie below

Figure 3 (A3): MAE for each bookmaker, elections 2010-2017, Con/Lab/Other basis
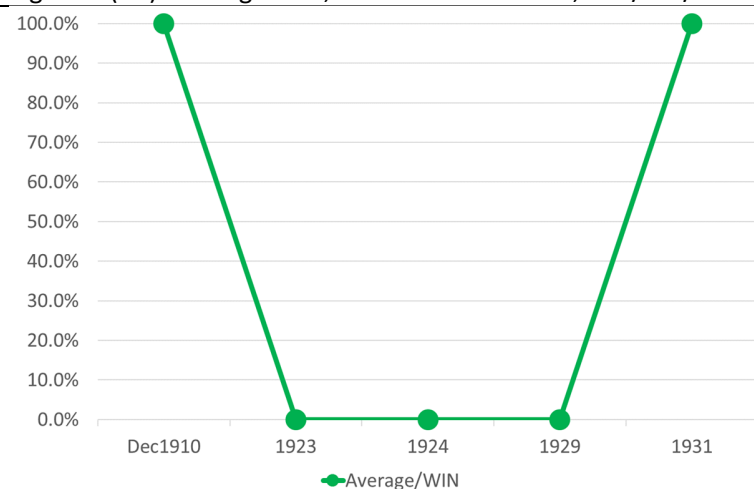

This graph depicts the MAE of the latest available odds on most seats (solid line) or overall majority (dashed line) offered by each of four bookies on the 2010-17 elections compared on a three-party-forced basis to the result. BF = Betfair, CL = Coral, LK = Ladbrokes, WH = WilliamHill.

Betfair Exchange is a betting exchange but the other three bookies are sportsbooks, so we'd expect Betfair to be more accurate given the faster pace of changing the odds. However that is not apparent from the data: Betfair was better in most seats in 2017, but is not consistently and noticeably better over elections. We must note that we did not consider inter-day variation and that may yield a different conclusion when examined.

WIN Index of the book

Now we consider WIN. Did our odds predict who would win?

Figure 7 (A4): Average WIN, elections 1910-1931, Con/Lab/Other basis

This graph depicts the average WIN of the latest available predicted majority (Dec1910, 1923, 1931) or seats (1924, 1929). The threshold is 50%. 100% means all correctly predicted the winner, 0% means none did.

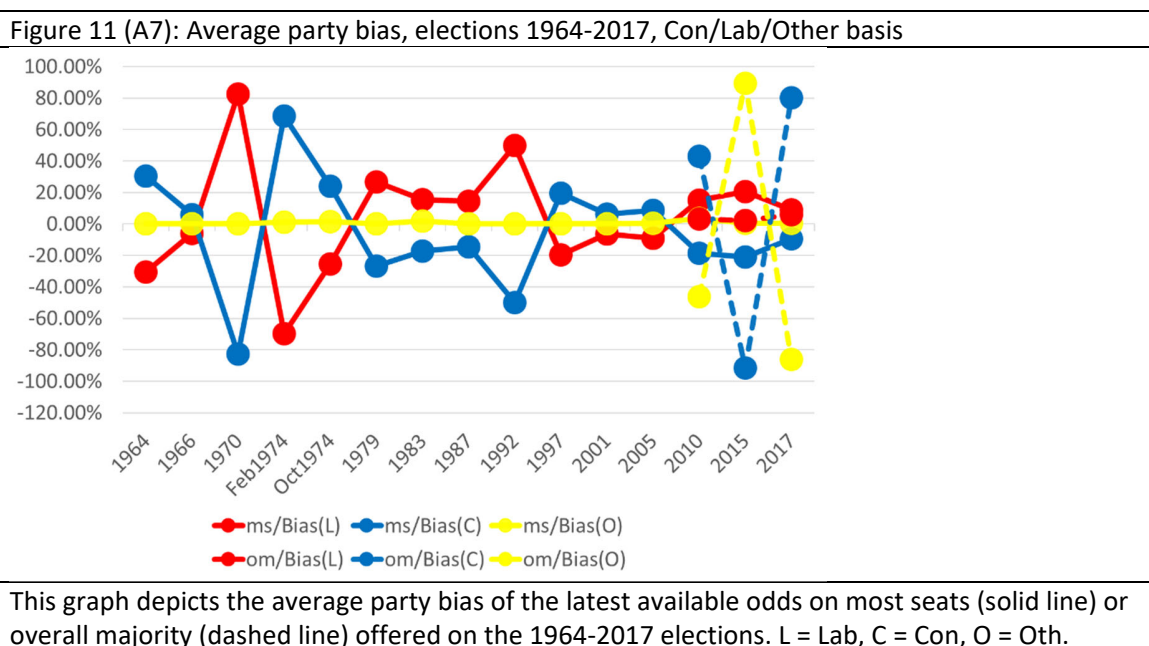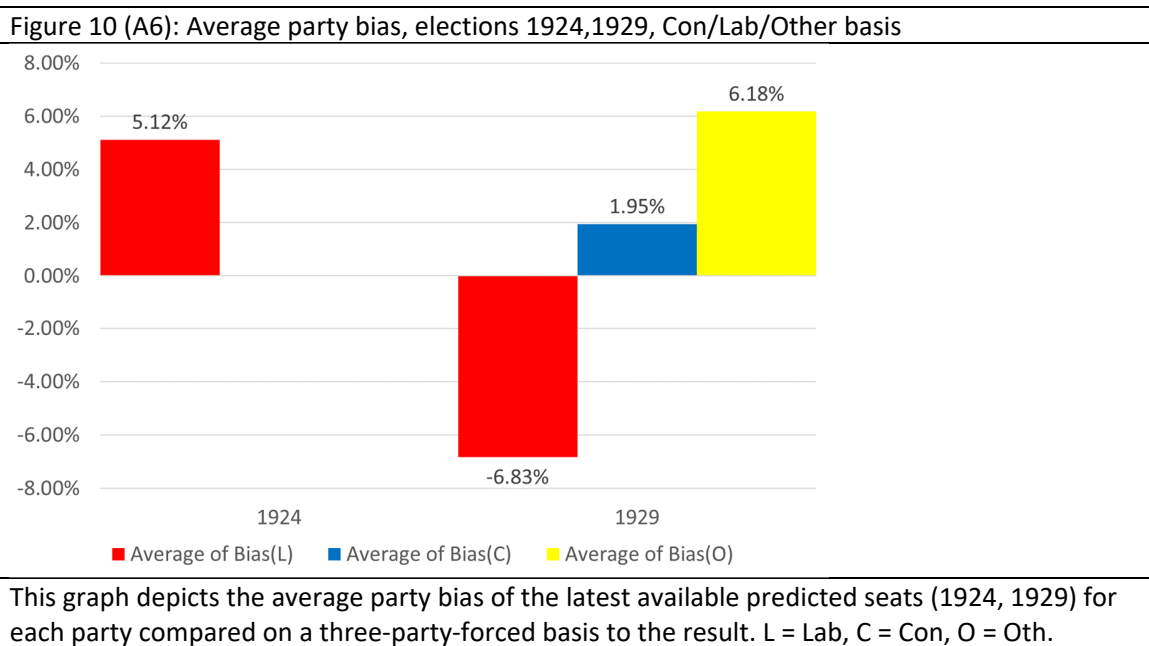Figure 8 (A5): Average WIN, elections 1964-2017, Con/Lab/Other basis
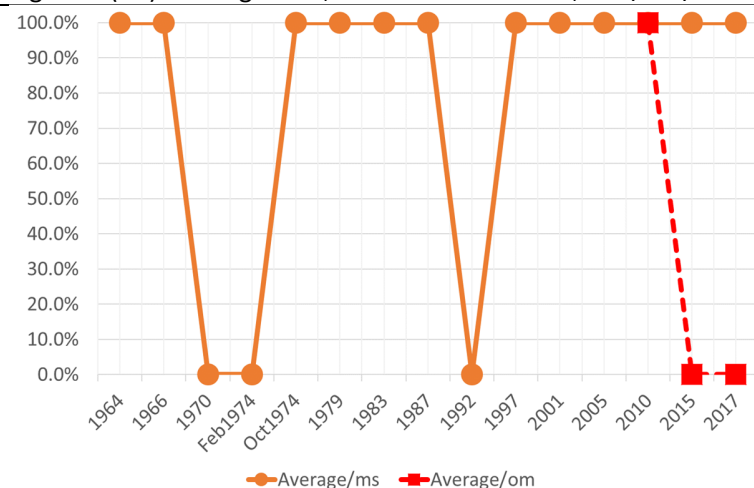
This graph depicts the average WIN of the latest available odds on most seats (solid line) or overall majority (dashed line) offered on the 1964-2017 elections. The threshold is 50%. 100% means all correctly predicted the winner, 0% means none did.

Looking at the WIN parameter this is considerably better, correctly predicting fourteen of the twenty elections considered in terms of most seats, although the overall majority is less good. Presumably this accounts for the good reputation odds enjoy among the public. It should be noted that they are tightly coupled: they succeed or fail together.
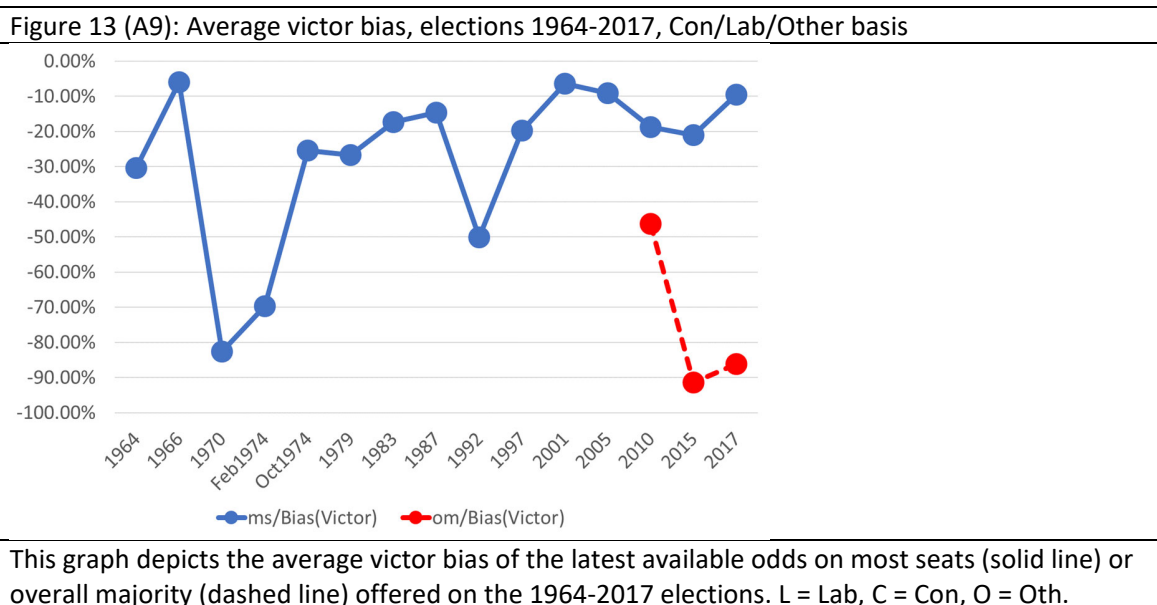
Bias of the individual odds

Now let's consider bias. Do we over- or underestimate a particular party?

Figure 10 (A6): Average party bias, elections 1924,1929, Con/Lab/Other basis

This graph depicts the average party bias of the latest available predicted seats (1924, 1929) for each party compared on a three-party-forced basis to the result. L = Lab, C = Con, O = Oth.

Figure 11 (A7): Average party bias, elections 1964-2017, Con/Lab/Other basis

This graph depicts the average party bias of the latest available odds on most seats (solid line) or overall majority (dashed line) offered on the 1964-2017 elections. L = Lab, C = Con, O = Oth.

Looking at the Bias parameters we note that, perhaps contrary to intuition, that the odds are not biased in any particular direction, although the bias for "Other" is considerably less than the bias for Labour or Conservative. Sometimes the errors are enormous: in 1970 the odds of most seats for the Labour party was 1/10, a raw probability of approx. 91%, but Labour lost the election. Nothing here is acceptable.

Now let's consider victor bias. Do we over- or underestimate the party that's going to win?

Figure 12 (A8): Average victor bias, election 1929, Con/Lab/Other basis

This graph depicts the average victor bias of the latest available predicted seats (1929) compared on a three-party-forced basis to the result.

Figure 13 (A9): Average victor bias, elections 1964-2017, Con/Lab/Other basis

This graph depicts the average victor bias of the latest available odds on most seats (solid line) or overall majority (dashed line) offered on the 1964-2017 elections. L = Lab, C = Con, O = Oth.

The odds consistently underestimate the probability of the eventual victor winning, sometimes hugely (1970 again) although that is unsurprising given favourite/longshot bias.

Calibration of the individual odds

We used MAE to judge the distance between the prediction and the actual outcome. Fair enough, but that's not the only way to judge probabilistic predictions. Instead of considering a single prediction deterministically, we can also consider several predictions frequentistically. Or to put it simply: are our 75% predictions right 75% of the time? This measurement is called "calibration"

The results are grouped into bins and the observed success rate for each bin is plotted on the y-axis against predicted success rate on the x-axis.[0908a] These plots are called "reliability diagrams", "calibration curves" or variants (not to be confused with the similar ROC curve), & look like this:

Figure 16 (A10): Reliability Diagrams, elections 1964-2017, Con/Lab/Other basis
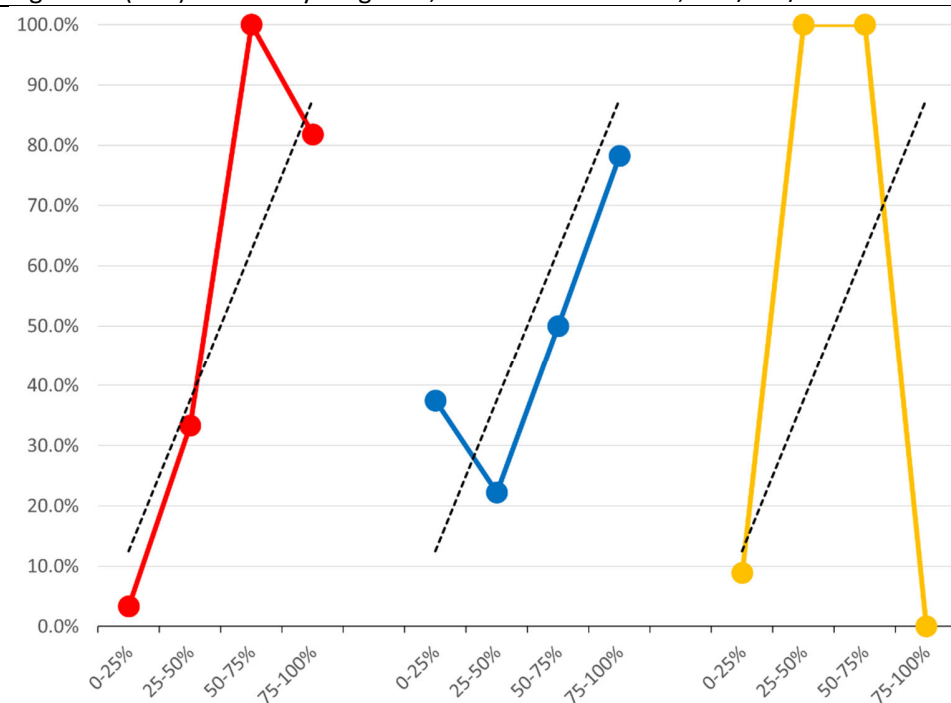
This graph plots the calibration pairs of the latest available odds on most seats and overall majority combined for the 1964-2017 elections. Red=Lab, Blue=Con, Yellow=Oth. Dotted line is the line of perfect calibration; points above/left of line are underestimates, points below/right are overestimates.

Labour and Conservative odds are pretty well calibrated, with some obvious exceptions (Lab odds in the 50-75% ranges are underestimates, as are Con in the 0-25% range). But the obvious problem is the "other" (Lib majority, no overall majority, etc): these odds are very unreliable. This unreliability may be a genuine phenomenon or it might be an artefact of the three-party-forced format. Another possible cause may be the fact that we combined the most seats and overall majority odds in order to get enough numbers. If we had enough data to consider each separately, we may have seen something different.

10. CURRENT UK ACCURACY

So we have considered historical gambling and looked at the accuracy over the years. So now we need to ask: how is gambling doing in the UK right now? This article was written in 2019, so we will define "current accuracy" as the accuracy of the 2015 and 2017 elections on the day before the election

2015

Firstly, let's look at 2015, The election was on 2015-05-07 and produced a Conservative overall majority

Table C1: accuracy of the 2015 books the day before the election. * = "imputed from previous days"

Book	Book	BF/ms	CL/ms	LK/ms	WH/ms	BF/om	CL/om	LK/om	WH/om
	Tstamp	06/05/2015 01:06	06/05/2015 01:06	06/05/2015 01:06	06/05/2015 01:06	06/05/2015 23:59	06/05/2015 00:47	06/05/2015 23:59	06/05/2015 23:59
Raw odds	Lab	_7/2	_4/1	_7/2	_7/2	_94/1	_40/1*	_33/1*	_40/1
	Con	_2/9	_1/6	_1/5	_1/5	_76/5	_7/1*	_10/1	_9/1
	LD	_949/1	_1000/1	_500/1	_1000/1		_1000/1*	_750/1*	_500/1*
	UKI		_250/1	_250/1	_200/1		_250/1*	_250/1*	_200/1*
	GRE		_1000/1	_1000/1		_1000/1*		_1000/1*	
	AOP	_807/1				_845/1			
	NOM					_1/16*	_1/16	_1/16	_1/16
Odds (3pf), overround removed	Lab	0.213	0.188	0.209	0.209	0.010	0.022	0.028	0.023
	Con	0.785	0.806	0.784	0.785	0.061	0.114	0.085	0.093
	OTH	0.002	0.006	0.007	0.006	0.929	0.864	0.887	0.884
Metrics (3pf)	MAE	0.144	0.129	0.144	0.143	0.626	0.591	0.610	0.605
	Precision	0.031	0.024	0.030	0.030	0.582	0.510	0.542	0.535
	Bias(L)	0.213	0.188	0.209	0.209	0.010	0.022	0.028	0.023
	Bias(C)	-0.215	-0.194	-0.216	-0.215	-0.939	-0.886	-0.915	-0.907
	Bias(O)	0.002	0.006	0.007	0.006	0.929	0.864	0.887	0.884

2017

Now let's look at 2017, The election was on 2017-06-08 and produced a hung Parliament, with the Conservatives on most seats but without an overall majority

Table C2: accuracy of the 2017 books the day before the election

| Book	Book	BF/ms	CL/ms	LK/ms	WH/ms	BF/om	CL/om	LK/om	WH/om
	Tstamp	07/06/2017 23:59	07/06/2017 23:59	07/06/2017 23:59	07/06/2017 23:59	07/06/2017 23:59	07/06/2017 23:59	07/06/2017 23:59	07/06/2017 23:59
Raw odds	Lab	_88/5	_8/1	_8/1	_8/1	_29/1	_16/1	_16/1	_8/1
	Con	_1/20	_1/20	_1/20	_1/20	_1/7	_1/6	_1/6	_1/6
	LD	_949/1	_500/1	_500/1	_500/1		_500/1	_500/1	_250/1
	UKI	_949/1	_1000/1	_1000/1	_1000/1		_1000/1	_1000/1	_500/1
	GRE		_1000/1	_1000/1	_1000/1		_1000/1	_1000/1	
	AOP	_949/1				_949/1			
	NOM					_41/5	_5/1	_11/2	_11/2
Odds (3pf), overround removed	Lab	0.038	0.072	0.072	0.072	0.033	0.054	0.055	0.098
	Con	0.002	0.007	0.007	0.007	0.859	0.789	0.798	0.760
	OTH	0.053	0.104	0.104	0.104	0.108	0.157	0.147	0.142
Metrics (3pf)	MAE	-0.056	-0.108	-0.108	-0.108	0.595	0.562	0.569	0.572
	Precision	0.003	0.004	0.004	0.004	0.384	0.334	0.342	0.331
	Bias(L)	0.038	0.072	0.072	0.072	0.033	0.054	0.055	0.098
	Bias(C)	0.002	0.007	0.007	0.007	0.859	0.789	0.798	0.760
	Bias(O)	0.053	0.104	0.104	0.104	-0.892	-0.843	-0.853	-0.858

11.   CURRENT UK ACCURACY BY TIME TO POLL

Let's now consider accuracy within an election: do books become more accurate as the election campaign progresses? The belief amongst gamblers that it does underlies the betting technique known as "value betting": if the odds on an event are wrong then they will converge to the real odds as the date of the election gets closer. Do books exhibit this behavior? Do they converge to zero as time to election reduces?

Figure 17 (P1): MAE of most seats odds over final 30 days, elections 2015-2017, Con/Lab/Other basis
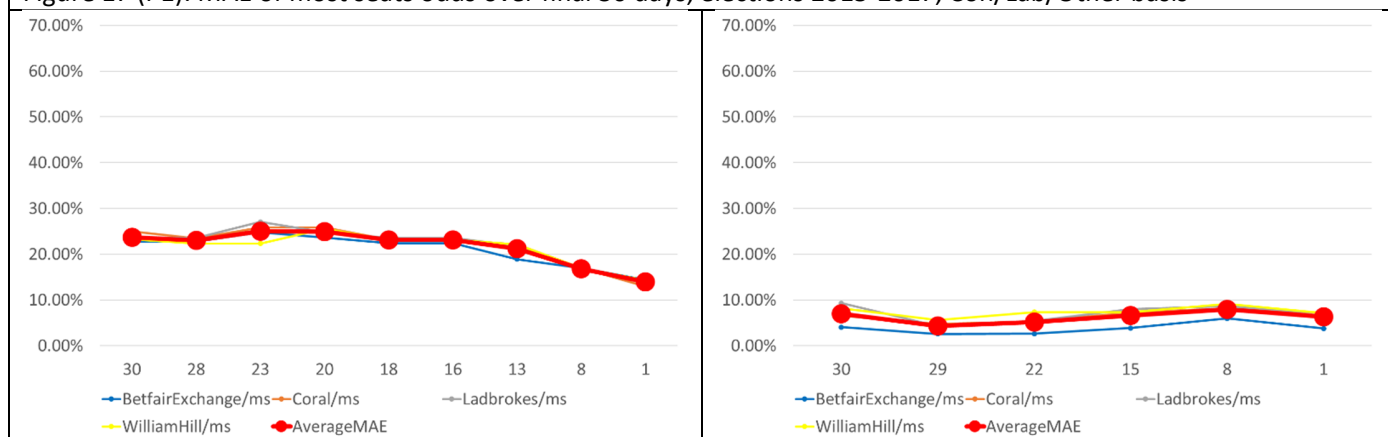
These graphs depict the MAE for latest available odds offered on most seats by each bookie over the 30 days prior to election day. Blue=Betfair, Brown=Coral, Grey=Ladbrokes, Yellow=WilliamHill, Thick Red=average.

Figure 18 (P2): MAE of overall majority odds over final 30 days, elections 2015-2017, Con/Lab/Other basis
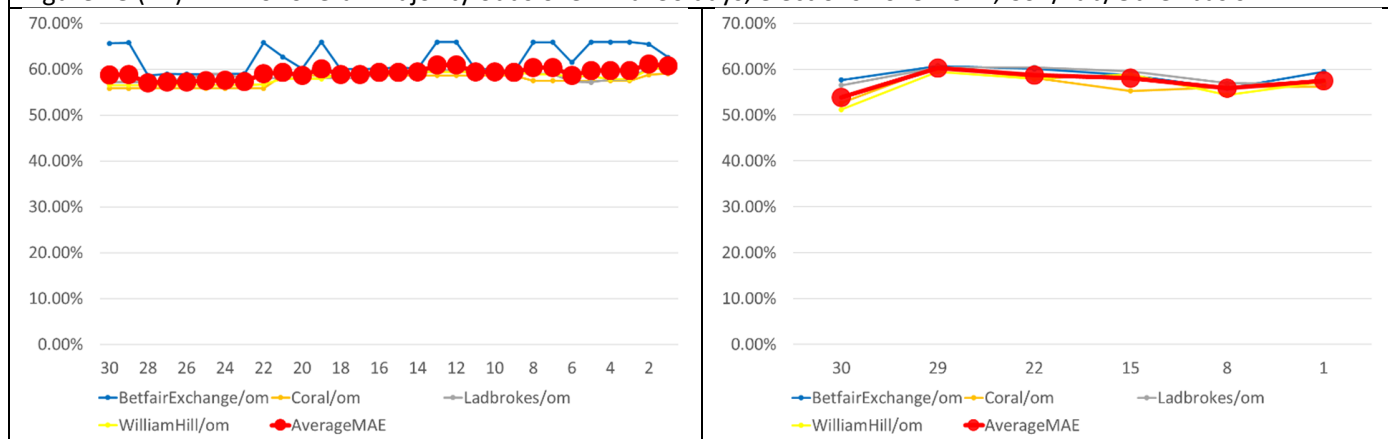
These graphs depict the MAE for latest available odds offered on overall majority by each bookie over the 30 days prior to election day. Blue=Betfair, Brown=Coral, Grey=Ladbrokes, Yellow=WilliamHill, Thick Red=average.

The "most seats" books exhibited this behavior, with a MAE either descending towards zero as the campaign proceeded or a low MAE at the start. But in both elections the "overall majority" had a MAE that was far too high and got slightly worse as time went on. Again, this may be a genuine phenomenon or it might be an artefact of the three-party-forced format, or a quirk of 2015 and 2017.

12. PUBLIC PERCEPTION

The general public and the media believe that gambling odds are the best indicator of electoral outcomes. The less-informed will cite a belief that odds are more reliable because people are betting their own money. The better-informed state that a perfect market and information will lead to an optimal solution. Phrases like "their own money", "wisdom of crowds" "smart money" will be bandied about. Alex Threlfall of Reuters interviewed Mike Robb of Betfair and in his preamble he said this:

> "…No opinion polls today of course but gamblers have long proved more reliable forecasters of election results than the pollsters, so where is the smart money going? Mike Robb of Betfair joins me now: Mike, thanks for coming in. BreakingViews, Reuters BreakingViews put out a piece this morning saying gamblers have only got it wrong in the US twice in a hundred-and-forty years…"
> -Alex Threlfall, Reuters, May 6 2010[0814c]. See [0814d] for the article mentioned.

Maxim Lott, a supervising producer of the John Stossel show on Fox, was similarly convinced, saying:

> "…Studies find that political prediction (betting) markets tend to be better at predicting elections than polls. Some reasons: Bettors take into account important factors besides polls. Unlike pundits, bettors put their money where their mouths are. People involved in the event might trade before news breaks publicly. The 'wisdom of crowds'…"
> - Election Betting Odds FAQ[0908b] discussing Betfair Exchange and PredictIt.

Bookmakers are adamant that their odds are good predictors. Consider the following statements from Ladbrokes and Coral personnel:

> "…The odds are arguably the most accurate predictor of what will happen on the 7th May [2015], as they are driven by hard, often informed, cash being invested, then our traders add in other factors like opinion polls to finalise those price …"
> - Simon Clare, Coral, April 15 2015[0814a]

> "…In the European elections…our odds were a better predictor of what the final scores were going to be for the parties than the polls were, and I suspect we'll see something similar at this election…It's always good to pay attention to where the money's going: that discipline of having to put cold, hard cash down is often a very good predictor of what the result's going to be…"
> - Matthew Shaddick, Ladbrokes, Jan 5 2015[0814b]

Professional gamblers hold a different view. It is known that gambling markets can be manipulated: Clement Freud did it routinely and the infamous "Trader B" spent seven million dollars keeping Romney's odds short on Intrade[0819e] for POTUS 2012. Indeed, whilst this article was being written the 2019 Conservative Party leadership election was held and Andrea Leadsom's odds were consistently shorter than her chances warranted, so to explain this a single individual nicknamed the "Leadsom Whale" willing to spend *à l'outrance* was hypothesized.[0819d] So the odds on an election can be manipulated by a deliberate actor.

But even if the odds of the event are unmanpulated, that doesn't stop them being wrong. In his book[0505d] Mike Smithson of PoliticalBetting notes the Dunfermline by-election of 9 February 2006. This election is notable because the odds pointed to the wrong winner right up to the very last moment: the odds showed Labour in the lead at 12:29am on the 10th, but a few seconds later the Liberal

Democrat candidate was declared the winner at 12:32am.[0505d]. Mr Smithson has this to say about the predictiveness of odds

> "…One thing that really annoys me is when people start suggesting that betting prices are the best guide to what is going to happen. If this were the case then favourites would always win. They don't. …I don't bet to provide a prediction tool for journalists who can't be arsed. I bet to try to win money…Betting prices are NOT a good indicator of political outcomes…"
>
> - Mike Smithson, PoliticalBetting.com, April 23 2015[0814e]

Academic work has focused on other forms of political betting such as prediction indices but despite their general belief academics such as Professors Laura Beers[0814g] and Leighton Vaughn Williams acknowledge that sometimes betting markets fail. Professor Vaughan Williams said in his written evidence to the House of Lords Select Committee on Political Polling and Digital Media:

> "…In summary, the overwhelming consensus of evidence prior to the 2015 UK General Election pointed to the success of political betting markets in predicting the outcome of elections. In contrast, the 2015 UK General Election, the 2016 EU referendum in the UK, the 2016 US presidential election and the 2017 UK election, all produced results that were a shock to the great majority of pollsters as well as to the betting markets. In each case, the longshot outcome (Conservative overall majority, Brexit, Trump, No overall majority) prevailed…"
>
> - Professor Leighton Vaughn Williams, Jan 16 2018[0814f]

13. CONCLUSIONS

The public believe that gambling odds are the best predictor. But being the best predictor is not the same as being a good one. Gambling odds, whether private books, majorities betting, sportsbook or betting exchanges can and do fail. This is not always the case, but in many recent cases they have failed, sometimes spectacularly. Academic work has focused on other forms of political betting such as prediction indices but even so some academics noticed that sometimes gambling odds fail and theories have been put forward to explain this recent underperformance.

Betting odds are simply the consensus of the betting community, weighted by wealth and risk appetite. Bookmakers offer a quick way to formulate, track and update that consensus but gamblers are not omniscient and, like every other human being, they make decisions based on their knowledge of the present and the past. But that is not sufficient for perfect prediction. As a report to the Market Research Society once said, "any attempt to predict the future depends on it resembling the past"[0423l]. In changeable times, when the past and present is no longer a guide to the future, people will fail and the odds will fail with them.

14. VERDICT

In the 5 general elections for which we have figures from Dec1910 to 1931, the odds were unacceptable for the winner of the election on 3 occasions: 1923, 1924, 1929. The threshold was at least fifty percent of the final odds predicting the winner of the election

In the 5 general elections for which we have figures from Dec1910 to 1931, the odds were unacceptable for winner probability on 4 occasions: 1923, 1924, 1929, 1931. The threshold was a mean absolute error MAE of 2.7% or less on a three-party-forced basis: Con/Lab/Other.

In the 18 general elections for which we have figures from 1964 to 2017 (fifteen most seats, three overall majority), the odds were unacceptable for the winner of the election on 5 occasions: 1970, Feb1974, 1992, and 2015 (om) and 2017 (om). The threshold was at least fifty percent of the final odds predicting the winner of the majority vote. Note that Coral maintain that they did offer odds in 1992 favoring Conservatives, but we could not obtain a contemporaneous source from the British Newspaper Archive.

In the 18 general elections for which we have figures from 1964 to 2017 (fifteen most seats, three overall majority), the odds were unacceptable for the winner probability on all occasions. The threshold was a mean absolute error MAE of 2.7% or less on a three-party-forced basis: Con/Lab/Other.

In the 52 books on a Lab/Con/Oth basis without overround for which we have figures from 1964 to 2017 (most seats and overall majority combined), the ungrouped Brier Score is 0.144 and the Brier Score of the books was 0.431. The Lab and Con odds are reasonably well calibrated with some notable exceptions, but the "other" odds are not well-calibrated.

Odds are good at predicting the winner on a binary basis, but that is all. The odds are well-calibrated but using a single odds or book as a probability predictor is not necessarily accurate.

15.   ACKNOWLEDGEMENTS

16.   NOTES AND CAVEATS

Notes and caveats on books versus odds

Many of the calculations in this article use a book – the set of all odds on a given election offered by a bookie at a given moment – as the basic unit of gambling, not an odds. Additionally, the odds will have the overround removed so that the total probability of the book adds up to one. Similar calculations on an individual odds with its overround intact may yield different results.

Notes and caveats on the Brier Score

Sources may differ on the formula used to calculate the Brier Score. Most examples of the Brier Score use the number of possible outcomes as the divisor, and using that method the Brier Score equates to the Mean Squared Error and hence varies from 0 to 1. But we also used the original Brier Score as defined by Brier in his 1950 paper[0602p], and original that uses the number of books as the divisor, not the number of possible outcomes. Using that method the maximum score may exceed 1, as Brier demonstrates.[0602p]

Notes and caveats on the 1992 General Election

Coral insist[0814a] that they made the Conservatives the favourites for the 1992 General Election. This claim may be entirely true. But the British Newspaper Archive makes no mention of it, and contemporaneous newspaper reports state instead that that Labour were favorites with the big bookmakers up to the last day,[0712z] then more money came in on election day[0712z][0712y] and the lead changed hands[0712z] before ending up as a dead heat.[0712z] So although Coral's claim may be true, without a contemporaneous timestamped source we cannot use it.

Notes and caveats on calibration

"Calibration" is a measure of the success of probabilistic predictions: for example, if you say that there is a thirty percent change that party x will win, and thirty percent of those predictions are correct, then the prediction is well-calibrated. But the question arises: is "well-calibrated" the same as "accurate"? Discussion sometimes elides "calibration" and "accuracy" together, and the picture is complicated by the use of the word "calibration" for non-probabilistic predictions. Phrases such as "accuracy and calibration of estimates"[0908d] or "confidence-accuracy miscalibration" [0908d] do not help. But other sources distinguish between "calibration" and "accuracy", either as separate concepts[0908f][0908c] or as set and subset[0908e]. Consequently this article will not treat "calibration" and "accuracy" as synonyms.

17.   OTHER

For the references and appendices and other notes and caveats, please see the accompanying article "Forecast Error: Supplement 2: Betting References"