BANDWAGON EFFECTS IN BRITISH ELECTIONS, 1885-1910

1. Introduction

Political parties seldom finish an election campaign – or, when polling is staggered over several days, the actual election – exactly as popular as they were when they started it. Whether the norm is a bandwagon effect when the early leader widens its lead, or an underdog effect whereby the initial runner-up closes in on the leader, continues to be disputed. It might, in any case, depend on the country concerned, on whether we are talking about opinion polls during a campaign or actual votes during a staggered election, or – if the election *is* staggered – on whether later polls are national runoffs (as in contemporary France) or first-round contests in districts that have decided to poll later (as in American presidential elections until 1844 and British general elections up to 1910).

Over the years, different studies have found bandwagon effects (McAllister and Studlar 1991; Mehrabian 1988; Nadeau, Niemi and Amato 1994), underdog effects (Fleitas 1971; Ceci and Kain 1982; Goot 2009) and no effect either way (West 1991). All of these studies focus on the effect of opinion polls upon the vote,[1] though there is a distinct U.S. literature examining how exit polls and television projections might influence voters in districts where the polls were still open. Here the main focus has been on overall turnout (Sudman 1986), with 'bandwagon' or 'underdog' effects occurring if one party's turnout dropped more than the other's once it was clear who

---

[1] Nadeau, Niemi and Amato (1994) is a partial exception to this in that they focus on voters' expectations over the outcome of the next election. However, one's expectations are likely to be significantly influenced by the polls.

was going to win. But again the evidence on bandwagons and underdogs is mixed (Delli Carpini 1984; Jackson 1983; Lott 2005.)

American primary elections have also been a fertile field for looking at the influence of earlier on later voters. Precisely because this influence is believed to exist, however, many states have brought their primaries forward in recent years, compressing the whole primary season. This is documented in Morton and Williams (2000), who also argue that the shorter the primary season, the slimmer are the chances of the original front-runner being overtaken by a rival. An underdog who achieves 'a dramatic finish in New Hampshire' may get enough money and media attention to go on to win the nomination, but not if the primary season is over almost before it starts.

British general elections before 1918 were staggered over a fortnight or more. Thus they also let us look at the effect of actual votes rather than polls or projections on subsequent voting behavior. And, whereas the course of American primaries may be driven, as Morton and Williams suggest, not just by reversals in campaign finance and media coverage, but also by tactical voting as particular candidates start or cease to look like potential winners, British elections in the 1885-1910 period were largely free of these varieties of noise. Once the voting began, the actual results normally drove any remnants of the campaign off the front pages. Most of the campaigners fell silent anyway. There is also less of a problem from the kind of tactical voting which might be mistaken for a bandwagon; the predominant modern form of tactical voting in Britain – voting for a center party because the left (right) candidate has no chance of keeping out the right (left) candidate – did not exist until after 1918. The 1885-1910 period has the further advantage that it featured no change in either voting qualifications or constituency boundaries.

We thus use the eight general elections between 1885 and 1910 to see if there was a general bandwagon or underdog effect as elections progressed. Existing studies assume that any such effect will be monotonic. This could be why they contradict one another as to which effect prevails. In our model, an initial bandwagon can go into reverse as voters not only learn different things but learn in different ways from successive rounds of results. (In particular, a strong bandwagon is more likely to be reversed than a weak one because of what it tells voters about who is driving it and why.) Furthermore our empirical analysis concludes that such a reversal was the typical pattern in British elections before a single polling day was instituted in 1918. On average, the bandwagon in favor of the party that eventually won peaked about halfway through the election and then declined. In theory, its decline could be due either to voters reining back some of their enthusiasm for the leading party – and our model indicates why they might – or to later polls involving voters less inclined to get on a bandwagon, any bandwagon, in the first place. Our results imply that both these factors counted.

Modern historians have rarely speculated on the subject. The principal instance is Lloyd (1968: 26), and even he goes no further than '.. the early results were generally believed to affect the later ones [but] it is not easy to demonstrate the existence of this bandwagon effect.' Nor did contemporary commentators often mention bandwagons or underdogs. Even when in 1895 *The Times* said that 'the moral effect of Sir William Harcourt's[2] defeat cannot fail to be very great' (Times, 15 July 1895: 9), it refrained from saying how the moral effect might translate into votes. Most of the references to possible bandwagons come after 1910, and in the context of the proposal to have all constituencies poll on the same day, finally adopted in 1918. Thus Herbert Nield MP

---

[2] Chancellor of the Exchequer in the Liberal government defeated at the 1895 election.

wanted to eliminate staggered polls because he believed that 'if the boroughs went in a certain direction and the results were announced it would have the effect of causing the counties to flow with whatever was the stream of victory' (Hansard, 5th series, vol. 99, col.1702, 26 November 1917). Henry Verney MP anticipated the disappearance of 'the wobbler' if all elections took place on the same day, as those electors who were on the look out to see 'how the cat jumped' would have to make up their own minds on political questions. (*The Times* 28th February 1914, p.12)

2. Literature

Some of the earliest psychological studies of human action focused on the influence of other people on an individual's behavior (Triplett 1898). As such, bandwagon effects and other similar phenomena are well documented in the behavioral psychology literature (See Crano 2000 for a review of the major literature on social influence). Initial studies by Sherif (1935, 1936), Hood and Sherif (1962), and Asch (1951, 1952, 1955, 1956) provide the foundations of social influence research, from which the two primary explanations of social influence were developed. The first explanation posits that bandwagon effects occur out of a desire to conform or a fear of the consequences of non-conformity. The second proposes that individuals derive information from others' choices and use this in conjunction with other information to make decisions. This is the 'rational bandwagon' approach. While the psychology literature as a whole is divided between the two approaches, though strongly favoring a bandwagon over an underdog effect, most economists and many political scientists in recent years have chosen the rational bandwagon approach. Here voters jump on and off the leading party's ship not out of a desire to conform or be different, but because they know that others have information that they lack, and so use the voting

or opinion poll figures to update their judgment as to whose government would provide them with the most utility. Banerjee (1992) looks at sequential decision-making in general. Different agents have different information, and those who decide later add to their own information by making inferences from earlier decisions. This 'information cascade' can lead to herd behaviour, i.e., a bandwagon. In Morton and Williams (2000) there is a two-stage election where voters are uncertain about the ideological locations of some of the candidates. Later voters use the early results both to make inferences about who stands where ideologically and to decide whether to replace sincere with strategic voting. This can lead either to bandwagon or underdog effects. Knight and Schiff (2010) model and test voting behavior in U.S. primaries. Voters estimate candidate quality by (1) a prior belief common to all voters, (2) a noisy signal common to all voters in each state, and (3) the voting by previous states. Candidates' reputations are boosted not when they have done well in previous states but when they have done surprisingly well, and Knight and Schiff conclude that John Kerry's unexpectedly good showing over Howard Dean in the early primaries helped clinch his nomination as Democratic presidential candidate in 2004.

Callander (2007) combines the 'conformist' and 'rational' approach to bandwagons in a model where later voters do read informational content into the early voting, but need to have a desire to conform as well to translate this into a bandwagon. He also shows that in staggered elections there might be an underdog effect. Suppose you don't want other voters to jump on a bandwagon, even one in favour of your own party, because you know that information is asymmetric (i.e. they have knowledge which you lack as to who would make the best government) and you want them to use it. Then your priority, as you vote, would be to stop a bandwagon getting started. You might even vote against your own preferred party in order to

achieve this. There could thus be an anti-bandwagon effect of a type that could not exist with simultaneous voting.

3.Model

Assume that voters are uniformly distributed along a left-right spectrum running from *y-1* to *y+1*. The ideological position of the median voter, *y*, fluctuates between elections. Voters and parties do not know its current value but they do know that it has a normal distribution around a mean of zero. In addition, all voters know the ideological positions that have been taken by the two parties. Each voter votes for the party whose government they believe would give them the greatest utility, where voter *i*'s utility from government by party *j* is:

$$U_{ij} = 2v_j - |\theta_i - \theta_j| \qquad (1),$$

where v_j is party *j*'s valence, i.e. its perceived competence irrespective of ideology (measured on whatever scale will give a coefficient of 2) and θ denotes position on the ideological spectrum. Each valence is normally distributed around a mean of zero.

There are two parties, *L* and *R*. They have to take an ideological position before the valences at the imminent election are known to them and, like the voters, they are ignorant of *y*'s current value but know that its expected value is zero. Symmetry thus

6

dictates that $\theta_L = -\theta_R \leq 0$.[3]

The election is held over three days, with each day's results known before the next day's polling begins. Voters are informed or uninformed. Informed voters know what the valences are, uninformed voters don't (though all voters know the parties' ideological locations). Let $v_R - v_L = x$, and let |x|, the absolute difference between the valences, determine the proportion of voters who are informed, *f|x|*, where *f'|x|>0* -- i.e., the stronger the signal that the parties differ in their competence, the more voters receive it. If there is no signal, no one can be receiving it, so that *f(0)=0,* and because *f|x|* cannot exceed unity although there is no limit on |x| itself, we assume that *f''|x|<0.*

It follows from (1) that informed voter *i* will vote for R if $\theta_i > -x$. Hence the proportion of informed votes that go to R is *0.5(x+y+1)* and so the total informed vote for R, as a fraction of the total informed and uninformed vote, is *0.5f|x|(x+y+1),* where *f|x|* is the proportion of voters who are informed.

Uninformed voters on the first day get no signal about party valences. They know only that *x* is normally distributed around a mean of zero, hence *E(x)_u=0* and uninformed voter *i* will vote for R if $\theta_i > 0$. The uninformed vote for R, as a fraction

of the total vote, is:

$0.5(1-f|x|)(y+1)$.

Hence,

$$S_{R1} = 0.5(xf \mid x \mid + y + 1) \qquad\qquad (2)$$

where S_{R1} is R's share of the total vote on day 1.

On the second day of voting, uninformed voters know a little more about x. S_{R1} is observable and, hence, they know the value of $xf|x|+y$. Assume initially that they also know the nature of $f|x|$, i.e., the relationship between the strength of the relative valence signal and its reach. They therefore know which combinations of x and y are consistent with the result they are observing. If, in addition, they know the pdf's of x and y, they can choose from these combinations, i.e., make joint maximum likelihood estimates $\hat{x}$ and $\hat{y}$. We assume that this is what they do.

On the second day, therefore, the proportion of uninformed votes won by party R will be $0.5(\hat{x} + y + 1)$ and it follows that R's share of the total vote will be:

$$S_{R2} = 0.5[xf \mid x \mid + \hat{x}(1 - f \mid x \mid) + y + 1] = S_{R1} + 0.5\hat{x}(1 - f \mid x \mid) \qquad\qquad (3)$$

So as they go to the polls on day 3, uninformed voters are uninformed no longer. They have seen S_{R1} and S_{R2} and so can solve equation (3) for $f|x|$, the only remaining unknown, and hence for x; x thus replaces $\hat{x}$ in equation (3) and, hence,

$$S_{R3} = 0.5[x + y + 1] = S_{R1} + 0.5x(1 - f \mid x \mid) . \qquad\qquad (4)$$

8

We now state the two key propositions. In their proof we assume, without loss of generality, that $x \geq 0$ (party R does not have the lower valence.) This enables us to replace $f|x|$ with $f(x)$, though we cannot do the same for $\hat{x}$, which could still be negative.

PROPOSITION 1: *The leader on the first day always gains further ground on day 2, i.e., every election starts with a bandwagon effect.*

Proof: We have assumed that uninformed voters on day 2 make a joint maximum likelihood estimation of x and y. The joint pdf for x and y is $P=p(x,y)=p(x)p(y)$ given that x and y are independent of each other.

Then $\frac{dP}{dx} = p'(x)p(y) + p(x)p'(y)\frac{dy}{dx}$,

where dy/dx represents the change in y needed to keep the voter's joint estimate of x and y consistent with the observed voting figures, after a unit change in x. From equation (2), $dy/dx = -\frac{d}{dx}(x(f(x))$

Hence, since $x \sim N(0, \sigma_x^2)$ and $y \sim N(0, \sigma_y^2)$,

$$\frac{dP}{dx} = \left(-\frac{x}{\sigma_x^2} - \frac{y}{\sigma_y^2} \cdot \frac{dy}{dx}\right)P$$

$$\therefore \frac{dP}{dx} = 0 \text{ when } y = \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{x}{-dy/dx} = \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{x}{\frac{d}{dx}(xf(x))} = \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{x}{f(x)+xf\prime(x)} \qquad (5)$$

i.e., $\hat{y}$ must have the same sign as $\hat{x}$.

Now suppose that party R leads on day 1. Then, from equation 2, $xf(x)+y>0$.

Therefore, $\hat{x}$ and $\hat{y}$ must both be positive. But according to equation 3,

$\hat{x} > 0 \Rightarrow S_{R2} > S_{R1}$. Hence, R extends its lead on day 2. (If party L leads on day 1, then $xf(x)+y<0 \Rightarrow \hat{x} < 0 \Rightarrow S_{R2} < S_{R1}$.)

The logic is that uninformed voters will interpret the party's initial lead as a sign that informed voters regard it highly, and will adjust their votes on the second day accordingly.

PROPOSITION 2 (i) *Whether the bandwagon continues on day 3 depends on whether*

$\hat{x} > x$.   (ii) *This in turn depends on whether* $y > \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{x}{f(x)+xf\prime(x)}$ *(a monotonically increasing function of x.)*

Proof: (i) From equations (3) and (4) , $S_{R3} - S_{R2} = 0.5(x - \hat{x})(1 - f(x))$

(ii)   Since, from equation (5), $\hat{y} = \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{\hat{x}}{-dy/dx}$

and since $\hat{x} > x \implies \hat{y} < y$

it follows that $\hat{x} > x \implies y > \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{\hat{x}}{-dy/dx} > \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{x}{-dy/dx} = \frac{\sigma_y^2}{\sigma_x^2} \cdot \frac{x}{f(x)+xf\prime(x)}$ (again from (5))

We now show that $\frac{x}{f(x)+xf\prime(x)}$ is monotonically increasing in x.

$\frac{d}{dx}\left(\frac{x}{f(x)+xf\prime(x)}\right) = \frac{f(x)-xf'(x)-x^2 f\prime\prime(x)}{[f(x)+xf'(x)]^2}$

But $f''(x) < 0$, while $\frac{d}{dx}[f(x) - xf'(x)] = -xf''x$, which is positive if $x > 0$.

$\therefore \frac{d}{dx}\left(\frac{x}{f(x)+xf\prime(x)}\right)$ is positive if $x > 0$. Given our assumption that $x \geq 0$, this makes

$\frac{x}{f(x)+xf\prime(x)}$ a monotonically increasing function of x. Thus $\hat{x} > x$ when y exceeds a

monotonically increasing function of x, i.e., when y is sufficiently high relative to x.

Hence any bandwagon for R will be reversed on the final day if $\hat{x} > x$, i.e. if y is sufficiently high relative to x ; any bandwagon for L will be reversed on the final day if $\hat{x} < x$, i.e., if y is sufficiently low in relation to x.

Thus, any party that owes its initial lead to ideology ends the election on a falling vote while a party that owes its initial lead to competence (valence) ends it on a rising vote. We have been assuming that $x \geq 0$. This does not necessarily mean that party R is the initial leader, but suppose that it is, and that it owes this lead mainly to ideology. Then y is large, x is small, and the swing to R on day 2 will be large (because small x means there are many uninformed voters, the only ones who change their minds). But day 2's success for R backfires the next morning when uninformed voters look at the results and say 'Most people must be as uncertain as I am about who would govern better, or they wouldn't be changing their votes in response to other people's votes on this scale. So the first day's Conservative lead says less in their favor than I thought and I'm less inclined to vote for them than I was.' By contrast, if day 1's lead had been down to valence there would have been quite a small bandwagon effect on day 2 because there are few uninformed voters. Day 3's uninformed voters now derive the leading party's large valence from this result, so that it does even better on the last day.

Before leaving the model, notice that some values of x, $\hat{x}$ and y will give S_R $\notin (0,1)$. In this case, the section of the electorate concerned will be unanimous. If on day 1 the uninformed voters are unanimous on the side of the current leader, then there can be no bandwagon and the results on days 2 and 3 will be the same as on day 1. Unanimity among informed voters, by contrast, will not stop uninformed ones joining or leaving bandwagons and the analysis is unchanged; while if the uninformed voters achieve unanimity only on day 2, then the bandwagon will be reversed iff

$\hat{x} > x$, just as in the non-unanimous case.[4]

4. Data

In attempting to identify the existence of a bandwagon there are a number of problems. First, if we see a leading party extending its lead during an election, or an election campaign, it does not necessarily mean that electors are flocking to the leading side because it is winning. Quite apart from the influence of events since the election began, it could be that the winning party's (presumably more convincing) message has had more time to get through to those who vote later. Secondly, it could be a case of tactical voting: some voters for instance might dislike hung parliaments and rally round the leading party, once they know which one it is, to give it an overall majority. Lastly, the central prediction of our model is that elections will start with a bandwagon – the initial leader will see its lead increased – but that this bandwagon might then accelerate, decelerate or even be reversed. For such complicated dynamics to take place, and be observable, it is necessary that the polls take place over a considerable time.

We regard all the above as reasons for choosing British general elections from

---

[4] Another point is that the model implicitly assumes that uninformed voters don't know the distribution of the electorate around y. If they do, then any $\hat{y} \neq y$ will be rejected by at least one uninformed voter as impossible, because there must be at least one voter outside the range $(\hat{y} - 1, \hat{y} + 1)$. Such voters will have to make an estimate of y, and therefore of x, that is closer to the true value, drawing day 2's results closer to the final results on day 3. If $\hat{x}<x$, this adjustment will strengthen the bandwagon (if it is in favor of party R) on day 2 but leaves it with less extra mileage to run on day 3. If $\hat{x}>x$, the bandwagon will now be weaker on day 2 and this will correspondingly weaken (but cannot wipe out) its reversal on day 3. We are grateful to one of the reviewers of the first draft of this article for suggesting both this line of inquiry, and also the point about possible unanimity discussed above.

1885 to 1910 as our object of study. We have already emphasized (section 1) the lack of opportunities either for tactical voting or for introducing new arguments once polling had started. Even to the minor extent that a few politicians tried to continue campaigning, less elaborate party organization and limited national publicity made it hard to introduce new arguments effectively and at short notice. The major exception in our sequence of elections was that of 1886 when, just after polling had begun, the veteran Liberal John Bright claimed that Gladstone had been converted to Irish Home Rule before the 1885 election and had deceived the public by not saying so until the voting was almost over. Both contemporaries and modern historians have seen this as contributing to the scale of the Liberal defeat in 1886 (Shannon 1999: 446) and, as will be seen later, our own analysis singles out 1886 as an election with one of the strongest initial swings towards the eventual winner. In general, however, British elections of 1885 to 1910 provide as neutral a setting for pure bandwagon or underdog effects as we are going to get.

Furthermore, the sheer length of UK elections before 1918 makes them an ideal candidate for our study. Each day, except Sundays, another batch of constituencies went to the polls, and each day many voters had access to the current state of play before they went out to vote. If, indeed, elections feature not just bandwagons, but bandwagons that can pick up speed, slow down, halt, go into reverse, or even be so far reversed that the initial winner is the ultimate loser, only a long drawn-out sequence of results is going to let us find out what did happen. While polling in the borough constituencies was usually over in the first week, the county contests stretched out day by day for over a fortnight. Small boroughs on average polled somewhat earlier than large multi-member ones, and Scottish and Welsh seats on average polled later than English ones. More generally, being far from London made

a difference: on average, each 100 miles of distance from London led to a delay of 2.5 days in the poll.

And the correlation between geography and earlier or later polling not only makes a promising basis for looking for bandwagon or underdog effects, but could shed light on what is driving them. A bandwagon that continues unabated suggests that geography as such is unimportant. But a bandwagon that slows down as it trundles out to remote localities might indicate that some regions are less inclined to react to a metropolitan trend. That might even be because they are less likely to have heard about it. The nineteenth century saw the beginning of the mass circulation daily newspaper, the primary source of information about the state of an election. But crucial to the role of the newspaper as means of mass communication were the development of the telegraph and the railways. The former ensured that the evening's results would appear in the provincial as well as the London press the next day. In 1889 there were 136 provincial morning or evening papers (excluding those devoted to a single subject, such as shipping or horse-racing) compared with 143 in 1900 and 128 in 1910 (Newspaper Press Directory, passim.) The minority of daily papers that omitted the previous night's national results did so not because of lack of access but because they limited themselves to local news. But the incompleteness of the national rail network -- some remote areas were still not well served – meant that election news would have taken longer to reach such areas.

Our analysis is run on those British constituencies[5] where contested elections took place between November 1885 and December 1910. Due to the limitations of swing in reflecting the dynamics of political change in a multiparty system, we

_____

[5] We exclude Irish and university constituencies as these constituencies did not reflect the political dynamics of the remainder of the United Kingdom.

exclude all constituency results in which a third party (i.e., party other than Liberal, Liberal Unionist or Conservative) came first or second.[6] Further, for the purposes of the analysis the Conservative party and the Liberal Unionist Party are treated as one party. We believe this is reasonable as the Liberal Unionists formed an informal coalition with the Conservative Party upon their formation in 1886, formed a formal coalition in 1895, and finally merged to form the Conservative and Unionist Party in 1912. Further, there are no seats in our sample in which a Conservative stood against a Liberal Unionist.

Data are used from the general elections of 1885, 1886, 1892, 1895, 1900, 1906, January 1910 and December 1910.[7] Election results were taken from the Society of Europe CD-ROM (Caramani 2000), while data regarding the day on which specific constituencies polled were constructed using Vincent and Stenton (1971), except for the election of December 1910, where we used *The Times*. Data on incumbency were taken from Craig (1974). Summary statistics for the results of each general election, along with statistics on the number of constituencies voting on each day, are included in the appendix.

---

[6] This exclusion results in 192 observations being dropped. Repeating our analysis treating Labour and the Liberals as one party (on the basis that a vote for either of these parties represents a vote against the Unionists) does not significantly alter our results (results not reported).

[7] Under the Septennial Act of 1715, no parliament could last more than seven years, but by convention the prime minister would ask for a dissolution in the sixth year of a parliament. But he might also call an election if he lost a vote of confidence, and this explains the premature election of 1886. In 1895, it was the incoming government (after the Liberals had lost a vote of confidence and resigned) that called the election. Both elections of 1910 came about during the Liberal government's clash with the House of Lords over Lloyd George's budget of 1909: in January they sought a mandate to push the budget through and in December to reform the Lords themselves.

5. Method and specification

Our dependent variable is 'Butler' swing (mean of one party's gain and the other's loss, as percentages of the total vote.) The use of swing, as opposed to the share of the vote gained by the winning party, has the significant advantage that it eliminates possible omitted variable bias due to constituency-specific effects that we are unable to observe. However, capturing a possible bandwagon effect on swing is complicated by the fact that swing is dependent not only on a party's performance in the current election, but also upon its performance at the previous election. Take two successive general elections and assume, without loss of generality, that the second of them is won by the Liberal party. Suppose that the Liberal share of the vote in constituency j in the second election is:

$$v_{j2} = V_2 + (v_j - V) + \beta_1 d_{j2} + \beta_2 d_{j2}^2 \tag{5}$$

where V_2 is the national share of the vote for the Liberals before any bandwagon or underdog effect gets under way, $(v_j - V)$ is the difference between the Liberals' share in seat j and their national share, again before any bandwagon starts to roll (we assume that $v_j - V$ doesn't change between the two elections), and d_{j2} is the number of days between the first national polling and the vote in seat j. The quadratic variable d_{j2}^2 is included because our model suggests that any bandwagon effect is unlikely to be linear in d. Coefficients β_1 and β_2 represent the strength of the bandwagon or underdog effect as follows:

$\beta_1 > 0$, $\beta_2 > 0$, strengthening bandwagon effect

$\beta_1 > 0$, $\beta_2 < 0$, weakening bandwagon effect

$\beta_1 < 0,\ \beta_2 < 0$, strengthening underdog effect

$\beta_1 < 0,\ \beta_2 > 0$, weakening underdog effect

Our model predicts a positive sign for β_1 but an ambiguous one for β_2. Now suppose that the previous election was also won by the Liberals, and suppose their share of the vote in seat j then was:

$$v_{j1} = V_1 + \left(v_j - V\right) + \beta_1 d_{j1} + \beta_2 {d_{j1}}^2$$

This means that the swing to the Liberals in seat j between the two elections is

$$SWING_{j2} = NATSWING_2 + \beta_1(d_{j2} - d_{j1}) + \beta_2({d_{j2}}^2 - {d_{j1}}^2)$$

where NATSWING means the national swing net of any bandwagon/underdog effect in either election.

If, however, the Liberals lost the first election, then we should write their share of the vote as

$v_{j1} = V_1 + \left(v_j - V\right) - \beta_1 d_{j1} - \beta_2 {d_{j1}}^2$. The Liberals' vote share will now fall (rise) as the days progress if there is a bandwagon (underdog) effect, so that

$$SWING_{j2} = NATSWING_2 + \beta_1(d_{j2} + d_{j1}) + \beta_1({d_{j2}}^2 + {d_{j1}}^2).$$

We thus write the general expression for the swing in seat j as

$$SWING_{j2} = NATSWING_2 + \beta_1(d_{j2} \pm d_{j1}) + \beta_2(d_{j2}^2 \pm d_{j1}^2) , \qquad\qquad (6)$$

where $\pm$ is plus or minus according to whether the party that won the second election lost or won the first one. What the formulation is saying is that, when a party wins two successive elections and a bandwagon effect exists, it will get its most favorable swing (in the second election) in those seats that polled early in the first election and late in the second one. Such seats have boarded a bandwagon. But if the winner of the second election had lost the first one, the biggest swing will be in seats that polled late in both elections, and have therefore climbed off one bandwagon onto an opposed one.

As all this implies, it is important for our estimations that constituencies do not poll in the same order in every election. If they did, then only those elections that defeated the winner of the previous election (so that we could use the fact that $d_{j2} + d_{j1} \neq 0$ even if $d_{j2} = d_{j1}$) would provide any test for bandwagon or underdog effects. In fact the average constituency had a modest but usable standard deviation of 1.98 days in its polling date across our run of elections.

In addition to the bandwagon terms we also include *Previous Majority, Incumbent Standing*, and *Third Party Vote. Previous Majority* measures how safe the seat is: it the size of the majority of whoever won the seat, as a percentage of the total votes cast. *Incumbent Standing* is complicated by the fact that many constituencies elected more than one MP. To attempt to model this, we opt for a trichotomous variable that takes the value of 1 if all the incumbents are standing, the value of 0.5 if at least one, but not all incumbents are standing, and 0 if no incumbents are standing. In single member districts *incumbent standing* is therefore a dummy variable. In addition to the above, a set of time dummies is included to capture the national swing

18

at each election. Furthermore, in our initial estimation, when we take each election separately, we include the control variable *DISTANCE* (from London) and the dummy variables *WALES*, *SCOTLAND, COUNTY* (a rural not a borough seat). A full list of variables and their definitions is presented in the Appendix.

Later we pool the elections in search of a generalized bandwagon or underdog effect, and add constituency fixed effects. This enables us to additionally eliminate time invariant factors that affect the size and direction of the swing and isolate within-constituency variation in the variables. Consequently we are no longer able -- and nor for our purpose do we any longer need -- to estimate the general effects on swing of the time invariant observed characteristics *DISTANCE, WALES*, *SCOTLAND, COUNTY*. But we can still interact these variables with the bandwagon variable to explore the magnitude of the bandwagon effect in different areas of Britain, and later we do so.

A Breusch-Pagan test (Breusch and Pagan 1979) of heteroscedasticity (p-value 0.000) revealed the presence of heteroscedasticity in our baseline regressions. consequently White 'robust' standard errors consequently are used throughout our analysis. We also test for AR(1) serial correlation between the error terms from successive elections (Wooldridge 2002). In all specifications utilized we are not able to reject the null hypothesis of no AR(1) serial correlation at the 5% level (p-values range from 0.063 to 0.718). [8]

---

[8] To ensure that serial correlation did not effect our results, a feasible generalized least squares procedure allowing for AR(1) errors was also used to test the robustness of our results. This did not alter the substance of our results (not reported).

6. Results

Table 1 shows the results for the 'bandwagon' coefficients β_1 and β_2 (p-values in brackets).

TABLE 1 HERE

Here, all but the last election have positive β_1 and negative β_2, suggesting an inverted U-shape relationship. But, although the number of significant results much exceeds what we would expect at random, eight out of 14 coefficients are not significant even at the 10% level. However, it could be that a bandwagon or an underdog effect will reveal itself with a larger number of observations. We therefore now pool all our data to see if a bandwagon pattern emerges across the sequence of elections taken as a whole. As explained in the previous section, we now use constituency fixed effects, which supersede the time-invariant control variables of the last regression.

6.1 Results with pooled data

The main results are as follows:

TABLE 2 HERE

Again – given that our model suggests that neither bandwagon nor underdog effects are likely to be a straightforward linear process – we try to capture the possibility that they might accelerate or go into reverse, and add a quadratic term. In column 1, the

coefficients on $(d_{j2} \pm d_{j1})$ and $(d_{j2}{}^2 \pm d_{j1}{}^2)$ are highly significant[9] and the latter coefficient is negative, indicating a bandwagon effect that eventually fades.

So when does the bandwagon peak? Equation (6) shows that the *swing* will peak, as a function of d_{j2}, when $d_{j2} = -\beta_1 / 2\beta_2 = 7.53$. Equation (5) shows that $\beta_1 / 2\beta_2$ also gives the date that maximizes the bandwagon in the second election.

If the bandwagon effect in elections between 1874 and 1910 did, on average, peak 7.53 days after the first polls, i.e., on day 8, the quadratic nature of our bandwagon curve means that it would have been eliminated another 7.53 days after this, i.e., on day 15. Given that the average election lasted around 15 days (there is no case where more than two results were outstanding after this time), we seem to have a rough parabola.

In terms of magnitude the figures tell us that, for example, a seat that polled on the first day of the previous election but on the eighth day of the current one (so that $(d_{j2} \pm d_{j1}) = 7$ regardless of whether the second election returns or defeats the incumbent) will have an additional 1.65% swing to the national winner as a result of the delay in the second poll. This is not an insignificant amount as 9.6% of seats in our sample had a majority of less than 3.3%.

6.2 Winner's bandwagon or gainer's bandwagon?

Until this point we have assumed that any bandwagon favors the eventual winner of an election. However it is possible that it is the party that is seen to be gaining seats,

---

[9] To test the robustness of our estimates to alternative samples we re-estimated our baseline model (equivalent to column 1 in Table 2) dropping each of the sample years in turn. Both $(d_{j2} \pm d_{j1})$ and $(d_{j2}{}^2 \pm d_{j1}{}^2)$ retain their significance at the 1% level in all samples (results not reported).

not the one seen to be heading for victory, that voters wish to join. (Magalhaes (2004), in his study of the Portugese elections of 2002, finds evidence for this latter case.) Column (2) in Table 2 therefore re-runs our baseline regression using swing to gainer as our dependent variable. Neither $(d_{j2} \pm d_{j1})$ nor $(d_{j2}^2 \pm d_{j1}^2)$ is significant. This is strong evidence that voters who loved a winner loved it for being a winner, not for being one on the favourable end of the swing.

6.3 Why did bandwagons decline?

We therefore return to the winner's bandwagon and its eventual decline. This decline does not necessarily indicate waning enthusiasm for the leading party. It could be that the seats which poll late are less susceptible to bandwagons. This would give us a slowing bandwagon (and thus a positive coefficient on $(d_{j2} \pm d_{j1})$ and a negative one on $(d_{j2}^2 \pm d_{j1}^2)$). But this is very different from the case where the incentive to jump on a bandwagon at first rises and then falls for any *given* voter. It could be that each member of the electorate is becoming ever keener on the leading party as the election progresses, but the people actually voting as the days go by are people progressively less inclined to get on bandwagons in the first place. [10] We have already discussed the possibility that voters in rural areas, a long way from London or outside England altogether might be less interested in, or even less cognisant of, the way the election

---

[10] A third possibility is that, as the eventual winner became more certain, turnout fell as voters downgraded their chances of making a difference (though this would reverse a bandwagon only if voters on the winning side did this disproportionately). We tested this by adding the change in each constituency's turnout since the previous election as a right-hand-side variable. If the slowing of the bandwagon effect was partly due to this kind of turnout effect, the coefficient on this new variable should be negative and significant, and it should also reduce the magnitude of the coefficient on the quadratic term (which would until now have been capturing the effect of the omitted variable). Since neither of these turned out to be the case, we conclude that the decline of the bandwagon was not due to voters' falling hopes of being pivotal.

was going nationally. Table 3, therefore, enters COUNTY, SCOTLAND, WALES and DISTANCE (from London) as interaction terms with the bandwagon variables $(d_{j2} \pm d_{j1})$ and $(d_{j2}^2 \pm d_{j1}^2)$ (DISTANCE is measured in units of 100 miles to eliminate superfluous zeros from its coefficient.)

TABLE 3 HERE

Being in a county rather than a borough seat does nothing to the bandwagon. But in Wales and Scotland the coefficients on both $(d_{j2} \pm d_{j1})$ and $(d_{j2}^2 \pm d_{j1}^2)$ are sharply reduced (though WALES* $(d_{j2}^2 \pm d_{j1}^2)$ falls just short of significance even at the 10% level). The coefficients on both SCOTLAND* $(d_{j2} \pm d_{j1})$ and WALES* $(d_{j2} \pm d_{j1})$ are opposite in sign to that on $(d_{j2} \pm d_{j1})$ and more than half its size; SCOTLAND* $(d_{j2}^2 \pm d_{j1}^2)$ and WALES* $(d_{j2}^2 \pm d_{j1}^2)$ bear the same relation to $(d_{j2}^2 \pm d_{j1}^2)$. Our estimate, therefore, is that the bandwagon effect was less than half as strong in Wales and Scotland as it was in England.

WALES and SCOTLAND encompass only a minority of constituencies, but DISTANCE covers them all. And in column 3 all of the key variables are significant and of the expected sign. A slowing bandwagon remains even when we control for the greater average distance from London of the later results. But the distance interactions are also significant, implying that our conjecture that distance brakes the bandwagon was justified.

Given the obvious collinearity between DISTANCE on the one hand and WALES and SCOTLAND on the other, it makes sense to take the interactions one stage further. Column 4 uses the variables of column 3 plus interactives of both

WALES and SCOTLAND with both DISTANCE* $(d_{j2} \pm d_{j1})$ and DISTANCE*

$(d_{j2}^{2} \pm d_{j1}^{2})$.   The result is that, compared with column 3, the coefficients on

DISTANCE* $(d_{j2} \pm d_{j1})$ and DISTANCE* $(d_{j2}^{2} \pm d_{j1}^{2})$ themselves are nearly doubled

and remain significant at the 1% and 5% levels respectively. The coefficients on the

four new interactive variables are insignificant, with the exception of

SCOTLAND*DISTANCE* $(d_{j2} \pm d_{j1})$. The import of this (positive) coefficient is

that being in Scotland reduces the effect of each 100 miles of distance from London

by about 42% (0.135 / 0.318). So a Scottish town 300 miles from London would, on

average, have as strong a bandwagon as an English one a mere 174 miles away. If

Scotland rode on a weaker bandwagon than England, that was due to its remoteness.

Being in Scotland 'as such' actually pushed up the bandwagon effect.[11]

But perhaps the main significance of column 4 is that the results of column 3 are

robust to the addition of new variables. The swing to the winner (and hence the

bandwagon effect) now peaks after 8.08 days, fractionally earlier than column 3's

estimate of 8.13, but later than our original estimate of 7.53. It still declines in the

second half of the election, even after we control for distance from London. Yet

distance from London significantly weakens the bandwagon too. Do we have a

bandwagon that slows as voters reconsider whether to join it, or a bandwagon that

slows because a different type of voter is now voting? The answer appears to be

'both.'

---

[11] Perhaps this was because the Scottish population was so concentrated around the Edinburgh and
Glasgow areas. In 1911, 55% of Scots lived in what is now the Strathclyde and Lothian region,
compared with the 20% of English who lived in the London area.

7. Conclusions

Our study examines British general elections between 1885 and 1910, building upon previous research that looks at the influence of opinion and exit polls on the way in which people vote. We exploit the fact that elections took place over a period of two weeks or more to examine the possibility of bandwagon effects as a result of previous votes as opposed to opinion polls. We find a significant overall bandwagon effect in British elections between 1885 and 1910. This seems to accord with Lloyd (1968) and Rhodes James (1963) whose historical analyses of British general elections in this period hint at the possibility of a bandwagon effect, as well as confirming the suspicions of the small number of commentators and parliamentarians who at the time theorized about this possibility.

Our analysis, however, suggests that the bandwagon effect was not monotonic, but instead had a tendency to peak about half-way through elections and then decline, but not to go into reverse, let alone to the extent of defeating the early leader. The eventual slowing of the bandwagon was caused *both* by voters' waning enthusiasm for the winning party *and* by the fact that later polls featured voters less prone to climb onto bandwagons in the first place.

Our analysis also demonstrates that location had an important influence on the extent of the bandwagon effect and that it was much weaker in Wales and Scotland, but comparable to other areas a long way from London. Scotland's bandwagon was actually stronger than its geographical position implies. We found no bandwagon effect in favour of the gainer (the party that gained seats at the election) as opposed to the winner (the party that won the most seats, whether it was gaining seats or losing them).

In the elections we have looked at, campaigning largely ceased once the actual polling began, suggesting that the way in which our voters learned from previous votes may differ from the way that modern voters use opinion polls. In their study of the American presidential election of 1988, Gelman and King (1994) imply that early election figures might carry some credibility for later voters that mere opinion polls lack. Their thesis is that an election campaign is a process of progressive enlightenment in which the reasons for voting for the parties are revealed as the campaign debate unfolds. This indeed is their explanation of the fact that the objective economic and political situation before the campaign starts usually is a better predictor of the outcome than most of the polls taken while the campaign is on. In the elections we have looked at, because campaigning largely ceased once the actual polling begun, we can regard the electorate, even on the first day of the poll, as being as enlightened as it ever would be, except for the fact that less informed voters had yet to read the signals from the actual results as they came in. To this extent, the way voters learned from actual results in the nineteenth century is distinct from the way they learned – if they did -- from opinion polls in the twentieth.

APPENDIX 1: Variable Definitions

Swing to Winner: The swing to the winner is calculated as follows:

(Δ Winner's percentage share of the vote – Δ Loser's percentage share of the vote)/2

When there was more than one candidate for a party standing, for example in a two member constituency, the total vote for all candidates of each party was used.

Swing to Gainer As above, with winner and loser replaced by party gaining ground and party losing ground.

Incumbent: A trichotomous variable that takes the value of 1 if all of the incumbents are standing for re-election, 0.5 if one but not all incumbents are standing for re-election and 0 if no incumbents are standing for re-election.

Majority: The percentage point lead in terms of the total vote in a constituency that the winning party has over the second place party. In two-member constituencies, majority is the percentage point lead of the lowest-polling successful candidate over the highest-polling unsuccessful candidate of a different party.

Scotland (Wales): A dummy variable that takes a value of one if the constituency is in Scotland (Wales) and zero otherwise.

County A dummy variable that takes a value of one if the constituency is a County constituency (as opposed to a borough) and zero otherwise.

Third Party Vote: The percentage of the total vote in a particular constituency that is not obtained by either the Conservative party, Liberal Unionist Party or the Liberal Party.

Distance: The distance in miles as the crow flies using the postcode EC4A 2DY (Fleet Street, London) to the local council offices applicable for that borough or county. Distance as the crow flies was calculated using the website Free Map Tools

http://www.freemaptools.com/distance-between-uk-postcodes.htm (accessed 13/10/11). If no council offices exist we used the high street. For counties we chose the town after which the county seat is named. Thus for the county seat Devon Honiton, we use Honiton for the purposes of the distance calculation. Where county seats do not include a town in the name we used the town closest to the center of the constituency. For example, for the county seat Norfolk Northern we used the town of Aylsham.

APPENDIX 2: Summary Statistics

TABLE 4 HERE

TABLE 5 HERE

REFERENCES

Asch, S. E. (1951). Effects of group pressure on the modification and distortion of judgments. In H.Guetzkow (Ed.), *Groups, leadership, and men.* Pittsburgh, PA.: Carnegie Press.

Asch, S. E. (1952). *Social psychology*. Englewood Cliffs, NJ. :Prentice Hall.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 193, 31–35.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, 70.

Banerjee, A.V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics,* 107 (3), 797-818

Breuch, T.S., & Pagan, A.R. (1979). Simple test for heteroskedasticity and random coefficient variation. *Econometrica,* 47(5), 1287-94.

Bruter, M., Erikson, R. S., & Strauss, A. B. (2010). Uncertain candidates, valence, and the dynamics of candidate position-taking. *Public Choice,* 144(1-2), 153-168.

Callander, S. (2007). Bandwagons and Momentum in Sequential Voting, *Review of Economic Studies,* 74(3), 653-84

Caramani, D. (2000). *Elections in western Europe since 1815: electoral results by constituencies.* London: Macmillan

Ceci, S.J., & Kain, E.L. (1982). Jumping on the Bandwagon with the underdog: The Impact of Attitude Polls on Polling Behaviour, *Public Opinion Quarterly* 46(2), 228-42

Craig, F.W.S. (1974). *British parliamentary election results: 1885-1918.* London: Macmillan

Craig, F.W.S. (1977). *British parliamentary election results: 1832-1885.* London: Macmillan,

Crano, W.D. (2000). Milestones in the Psychological Analysis of Social Influence. *Group Dynamics,* 4(1), 68–80.

Delli Carpini, M.X. (1984). Scooping the Voters? The consequences of the networks' call of the 1980 presidential race. *Journal of Politics,* 46(3), 866-85

Fleitas, D.W. (1971). Bandwagon and Underdog Effects in Minimal-Information Elections, *American Political Science Review* 65(2), 434-8

Gelman, A., & King, G. (1994). Why are American presidential Campaign Polls So Variable When Votes Are So Predictable? *British Journal of Political Science,* 23(4), 409-51

Goot, M. (2009). Underdogs, Bandwagons or Incumbency? Party Support at the Beginning and End of Australian Election Campaigns, 1983-2007. *Australian Cultural History,* 28, 69-80

Groseclose, T. (2001). A model of candidate location when one candidate has a valence advantage. *American Journal of Political Science,* 45(4), 862-886.

Hood, W.R., & Sherif, M. (1962). Verbal report and judgment of an unstructured stimulus. *Journal of Psychology,* 54(1), 121-30.

Jackson, J.E. (1983). Election Night Voting and Voter Turnout. *American Journal of Political Science,* 27(4), 615-35

Knight, B. & Schiff, N. (2010). Momentum and Social Learning in Presidential Primaries. *Journal of Political Economy,* 118 (6), 1110-50.

Lloyd, T. (1968). *The General Election of 1880.* Oxford: Oxford University Press.

Lott, J.R. (2005). The impact of early media election calls on Republican voting rates in Florida's western Panhandle counties in 2000. *Public Choice,* 123(3-4), 349-61

Magalhaes, P.C. (2004). Exposure to Polls, Cognitive Mobilization, and Voting Behavior: the 2002 General Elections in Portugal. Research note, [online]. Available from: http://www.ics.ul.pt. [Accessed 18 May 2011].

McAllister, I., & Studlar, D.T. (1991). Bandwagon, Underdog, or Projection? Opinion Polls and Electoral Choice in Britain, 1979-1987. *Journal of Politics,* 53(3), 720-41.

Macdonald, S.E., & Rabinowitz, G. (1998). Solving the Paradox of Nonconvergence: Valence, Position and Direction in Democratic Politics. *Electoral Studies,* 17(3), 281-300

Mehrabian, A. (1998). Effects of Poll Reports on Voter Preferences. *Journal of Applied Social Psychology,* 28(1), 119-30.

Morton, R.B. and Williams, K.C. (2000). *Learning by Voting*. Ann Arbor: University of Michigan Press.

Nadeau, R., Niemi, R.G., & Amato, T. (1994). Expectations and Preferences in British General Elections. *American Political Science Review,* 88(2), 371-83.

Newspaper Press Directory (1846-1976). London: C.Mitchell.

Rhodes James, R. (1963). *Rosebery.* Weidenfeld and Nicolson, London.

Schofield, N. (2004). Equilibrium in spatial valence models of politics. *Journal of Theoretical Politics*, 16(4), 447-481.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology,* 27, 1–60.

Sherif, M. (1936). *The psychology of social norms*. New York: Harper & Row.

Sudman, S. (1986). Do Exit Polls Influence Voting Behavior?' *Public Opinion Quarterly,* 50(3), 331-9

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology,* 9(4), 507–533.

Vincent, J.R., & Stenton, M.M. (Eds.) (1971). *McCalmont's Parliamentary Poll Book of All Elections, 1832-1918.* Brighton : Harvester Press

West, D.M. (1991). Polling Effects in Election Campaigns. *Political Behavior,* 13(2), 151-63.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge MA: MIT Press.