Measuring Polling Bias in 2008: The Bradley Effect and Related Issues

Reuben Kline and Christopher Stout
University of California Irvine
Department of Political Science, and
Center for the Study of Democracy

April 27, 2009

Abstract

We put forth five criteria which should be met by a suitable measure of the Bradley effect, and assess the validity of nine measures of polling inaccuracy originally discussed by Mosteller (1948) and Martin, Traugott and Kennedy (2005) with respect to these five criteria. The only two measures which have heretofore been used in studies of the Bradley effect are equivalent on four of the five criteria (as is a third, newer, measure). In fact, these two measure are equivalent up to a constant of proportionality, but *if and only if*: 1) an exhaustive allocation of the undecided proportion of the sampled electorate has been made and 2) there is no third party support, in either the poll or the election. Using hypothetical examples and data from the 2008 Presidential election we demonstrate that one of these measures, used in Kline and Stout (2008) remains invariant to changes in the proportion of undecided voters and to the existence of third party support, while the other (used in Hopkins, forthcoming and Stromberg, 2008) can fluctuate dramatically depending on the degree to which there are undecided voters and support for third party candidates. As a result of this, we strongly recommend that scholars employ the former in any and all investigations into the Bradley effect and analogous phenomena.

Introduction: The Bradley Effect and Measures of Polling Inaccuracy

The historic nature of the 2008 election has led to a plethora of research projects analyzing several aspects of this unique election. Of particular interest to academics and non-academics alike was whether pre-election polls were accurately gauging Barack Obama's electoral support. As a result, there has been a renewed interest in the phenomenon known as the "Bradley Effect[1]". The Bradley Effect refers to a particular type of polling bias in which electoral support for an African-American candidate is over-estimated due to a tendency for white voters to misrepresent their true preferences (presumably for the white candidate) in order to obviate claims of racism or bigotry.

Although polls were quite accurate in measuring electoral support for McCain *and* Obama, the historic nature of this election generated a large and growing number of papers analyzing the Bradley Effect, several of which (Kline and Stout, 2008; Stromberg, 2008; and Hopkins, forthcoming) have taken a retrospective look at past African American candidacies. Though these studies have been quite sophisticated, they nonetheless reach inconsistent conclusions (Greenwald and Albertson, 2008; Hopkins, Forthcoming; Hugick and Zeglarski, 1993; Keeter and Samaranayake, 2007; Stout and Kline, 2008; Stromberg, 2008). These divergent results could have been caused by a number of factors, but one possible explanation is that the studies did not use a common measure of polling inaccuracy. While there have been numerous studies that have discussed the best measure for assessing polling inaccuracy for elections in general (Martin, Traugott and Kennedy, 2005; Mitofsky, 1998; Mosteller et al, 1949), there is no research that has discussed the advantages and disadvantages of the various existing

[1] Hopkins (forthcoming) employs the term "Wilder Effect" but the logic of the phenomenon is identical.

measures of polling error in an application such as the Bradley effect. As a result, those researching the Bradley Effect have taken for granted the measurement of the underlying phenomenon: polling bias or inaccuracy.

The goal of this paper is to identify the most appropriate measure for polling inaccuracy in the context of the Bradley effect (and other phenomena which involve polling inaccuracy attributable to a particular candidate characteristic). To find the most suitable measure for the "Bradley Effect" we first outline a set of five criteria for an appropriate measure of the Bradley Effect. Next, we examine the nine most frequently referenced methods (measures put forth by Mosteller, et al, 1949; and Martin, Traugott and Kennedy, 2005) for assessing polling inaccuracy in an election and evaluate their suitability as measures of polling bias arising from the Bradley effect or analogous phenomena.

After the initial assessment of the measures with respect to criteria one, two and three, we are left with three seemingly suitable measures (#2, #5, and #9). We then further analyze these three methods and determine which one best measures polling inaccuracy for our purposes, based on the fourth and fifth criteria. Using hypothetical data and data from the 2008 Presidential election, we show that #5 is the best measure of the Bradley Effect because it measures candidate specific polling support, is the most consistent (we will explain precisely what we mean by consistent below) of the three measures across many elections with varying characteristics and, moreover, is the most intuitive. Finally, using state-level data from the 2008 Presidential race, we compare and contrast #2 and #5, and find that #2 deviates from #5 in exactly the ways in which we

would expect given our preceding analysis. Overall, the results suggest that #5 is the preferred measure for assessing the "Bradley Effect"

Nine Measures of Polling Inaccuracy

Table 1 contains formulae, in as consistent notation as possible, for the nine most frequently referenced and discussed measures of polling inaccuracy. The first eight correspond exactly to the eight measures discussed in Mosteller et al (1949) and Mitofsky (1998). The ninth is a newly developed measure due to Martin, Traugott and Kennedy (2005). Once these are all rendered in a common notation, their individual properties as well as the relationships between them become more clear than is the case in the (often confusing) verbal descriptions offered in Mosteller et al (1949) and Mitofsky (1998). We now assess each of the measures based on the criteria above, taking them one at a time.

[Insert Table 1 about Here]

The Bradley Effect: Criteria for an Ideal Measure of Polling Accuracy

Given that multiple measures are available for measuring polling inaccuracy, we must assess their relative suitability with respect to polling inaccuracy for the specific purpose of assessing claims related to the Bradley effect. To this end, we outline five criteria for a measure appropriate for such an application. First, because the Bradley effect is not concerned with the ultimate winner of the election, but rather the polling-performance gap for an African-American candidate, we want a measure which is *not based on the winner* of an election but is, rather, somehow related to the margin between the candidates or their respective vote shares. This is illustrated by the two most commonly cited cases of the Bradley effect (including the eponymous one): Tom

Bradley's 1982 candidacy, in a race which he lost despite being projected to win by a large margin, and Douglas Wilder's 1989 candidacy, in an election which he won, but by a margin much slimmer than was projected by the polls.

Second, since analyses pertaining to the Bradley effect must involve estimating the effect of a particular personal characteristic shared by a group of candidates, a proper measure of it must be *candidate specific*, so that it isolates the inaccuracy with respect to that particular candidate or group of candidates. Therefore it must not be a measure of the general level of inaccuracy in the overall election, such as an average deviation across *all* the candidates.

Third, because the *proportion of undecided voters* is a conceptually important part of the theory behind the Bradley effect (Berinsky, 1998; Traugott and Price, 1992), a suitable measure must not require a post-hoc allocation of the undecideds. Berinsky (1999) shows that respondents who want to deceive pollsters about their voting intentions will often identify as undecided even if they have a preference for a black candidate's opponent. As we shall see below, a measure which is not identical with and without the undecided voters allocated has a great potential to obscure the cause held to be the driving behind the Bradley effect: socially desirable response bias.

Fourth, an appropriate measure method is one which allows us to best *compare the magnitude* of the Bradley Effect consistently across time, across candidates and *across many elections*. Consistency is important when measuring any type of polling accuracy. Inconsistent estimates make it difficult to make comparisons across large number of cases, which essentially precludes meaningful analysis across large numbers of cases. Fifth and finally, though this is a more subjective matter, we believe that a

proper measure must be intuitive and readily interpretable. Perhaps a good proxy for this criterion is the extent to which a measure is (implicitly or explicitly) used in discussions of polling inaccuracy, in both popular and academic settings.

Applying the Five Criteria to Nine Measures of Polling Inaccuracy

Criterion 1: The method is not chiefly concerned with the prediction of the eventual winner:

Method # 8 is ultimately a measure of the aggregated capacity of the individual state-level polls to correctly predict the winner in their respective states. As a result, #8 exclusively focuses on predicting Electoral College votes. The "Bradley Effect" is not only concerned with who ultimately wins the election, rather its focuses on whether polls were drastically overestimated support for the candidate of interest. Therefore, Method #8 is not a suitable measure for the Bradley Effect. All other methods are not primarily concerned with the ultimate winner in the election, but instead relate to the accuracy of the predicted relative vote shares or margins of victory.

Criterion 2: The method is candidate specific:

While many studies of the polling accuracy are meant to measure how well polls predict the election outcome, research on the Bradley Effect or analogous phenomena focuses more on how well the polls measure support for a particular candidate (generally a minority candidate). Methods #3, #4, #6, and #7 all are, in one way or another, measures of the general inaccuracy of polling in a given election, and thus are not candidate specific. Therefore, these five measures should not be considered for use in analyses of the Bradley effect, because they cannot measure a candidate-specific

inaccuracy. This is not to say that some or all of them are not useful in any context, but that they are not appropriate for any analysis of polling inaccuracy which stems from a particular characteristic of a candidate or group of candidates.[2]

Criterion 3: The Method does not require any allocation of the undecided voters:

We are now left with #1, #2, #5 and #9 as viable candidate measures. While #1 is a candidate-specific measure it has the disadvantage of systematically exaggerating the degree of polling inaccuracy if the undecided proportion of the poll is not exhaustively allocated, a shortcoming it shares with methods #3, #4, #6 and #7 (Crespi, 1988; Mitofsky, 1998).

The remaining measures—#2, #5 and #9—all meet our first three criteria: they do not measure the success of predicting the eventual winner; they are candidate-specific; and they have, by construction, eliminated the undecided proportion in the polls. As aforementioned, the formulation of the measures in Table 1 allows us a better view of the underlying relationships among the three constructs. For #2 we subtract the polled share of the two-party vote for A, our candidate of interest, from the actual share of the two-party vote. For #5 we calculate first the predicted margin of victory for the candidate of interest, and then we calculate the actual margin of victory. Thus this method is a difference between two margins of victory. #9 employs a log-odds ratio. In this case, it is the natural logarithm of the ratio of the odds of a victory for candidate A based on the poll (P_a/P_b), and the odds of candidate A's victory based on the election results (V_a/V_b).

[2] In fact #3, along with #5, has been used to assess polling inaccuracy in many studies including, *inter alia*, Mitofsky (1998), Malghaes (2005), Traugott (2001), and Panagakis (1999).

Criteron 4: The Method allows consistent comparisons across many elections, with differing characteristics:

Normalizing away the undecided proportion, as is done in #2 and #9, does not ensure, however, that the proportion of undecideds does not present a problem for comparison across elections. Because #2 and #9 contain normalizing divisors (P_a+P_b and V_a+V_b in the case of #2 and P_b and V_b in the case of #9), they are not as robust as #5 to the inclusion of and/or changes in several key election parameters: the proportion of undecided voters, the degree of third-party support, and the competitiveness of the race (as measured by the margin of victory predicted by the polls), even when all other relevant factors are held constant.

Because it considers only the differences between two margins of victory, #5's estimate of the polling error remains constant across cases irrespective of changes in the proportion of undecided voters or in the difference between the share of the major party vote in the pre-election poll and the final result. A one percent discrepancy between the poll and the final result is the same if fifty percent of the electorate was undecided or if two percent were undecided. The same holds for the margin of victory. Because #2 and #9 both contain normalizing divisors, their estimates of the polling error, on the other hand, are guaranteed to be robust *if and only if* the major party vote and poll shares are at unity (which implies an exhaustive allocation of the undecided voters and no third-party support). In fact, if it is the case that the major-party vote and poll shares are at unity, then it can be shown (as is done in Appendix 1) that the value of #5 is exactly twice that of #2.

Comparison of Measures with respect to *Criterion 4*: Hypothetical Data

To illustrate the impact of the existence of undecided voters in the poll, imagine a hypothetical scenario. In this scenario, assume that the population sampled in the poll consists of the very same people who vote in the election, and these are the only people who vote. Furthermore, in this idealized scenario, we are able to pair each respondent's poll response to her vote choice. Thus, in such a scenario, we can directly measure the degree of polling inaccuracy[3], so, which measure—#2, #5 or #9—would be preferred under these ideal conditions? (as a phenomenon of polling bias, not from the standpoint of individual-level motivations to misrepresent preferences)

In this case, a poll finds 40% support for candidate A, who happens to be black, and 45% support for candidate B, who is a white male. The poll also reports 15% undecided, and there are no third-party candidates. Assume that two-thirds of the undecided voters (10% of the vote in total) actually support candidate B but, out of concern that they may be perceived as racist, reported their voting intention as undecided. All other voters voted according to their intention in the poll, with the remaining 5% of undecideds—those that were genuinely undecided—being split evenly between the two candidates, thus the election outcome was 57.5% in favor of B and 42.5% for candidate A. Without allocating the undecided, the Bradley effect as measured by #5 is 10. We divide this number by 2, to see how much #2 deviates from this ideal proportion. #2 in this example yields a score of 4.6.

[3] Here we are merely attempting to find the best measure for polling inaccuracy, which is a necessary first step to determining the *existence* of the Bradley effect. The *explanation for* the Bradley effect, such as individual –level motivations, while interesting in their own right, are beyond the scope of the topic at hand.

Now consider the same candidates, but assume that instead of A polling 40%, he polls 25%, and candidate B polls 60% instead of 45%. The undecided proportion remains the same (15%) and the undecideds vote in the same manner and proportion as they did in the first example, thus the final vote is 72.5% for B 27.5%. We know by virtue of our omniscience that an additional two-thirds of the voters (10% in total) misrepresented their voting intention by claiming to be undecided when they actually preferred B to A. This is the same as in the first example, when #5 yielded a Bradley Effect of 10. Again, we see that #5 divided by two yields a Bradley Effect of 5. In this case, however, method #2 yields 1.9. This example is re-created below in Table 2A.

Thus, although the underlying phenomenon—the proportion of respondents misrepresenting their true preference—did not change, the effect reported by #2 and #9 changes drastically. Moreover, the degree of the Bradley Effect is underestimated in precisely the context in which many Black candidates run: uncompetitive elections.[4] The polling inaccuracy reported by #5 in this example, however, is invariant to these changes, giving a more consistent measure and thus one which is more suitable for comparisons across elections.

There has been some discussion in the literature of the proper way to allocate undecided voters (see Berinsky 1998 for a thoughtful and useful discussion), and such an allocation might be perfectly reasonable under some circumstances. Nonetheless, at least in the case of the Bradley Effect (or an analogous phenomenon which is in part related to the behavior of the undecided proportion of the sample), any attempt at allocating undecided voters is bound to obscure the true nature of the effect. Thus, making accurate

[4] Add in percent of Black candidates who ran in election that were decided by more than 10 points.

comparisons regarding the magnitude of the polling inaccuracy across elections using either #2 or #9 is made more difficult to the extent that there are undecided voters in the sample of the poll in question.

[INSERT TABLE 2 ABOUT HERE]

Though in these examples we could exhaustively allocate the undecideds and thus derive a consistent measure using #2 as well, in doing so we will have largely obscured the very effect we are attempting to measure. In fact, as Hugick and Zeglarski (year??) point out, in a race involving a black Democrat and a white Republican, the undecided voters are likely to break disproportionately for the white Republican candidate. Assuming this is true, this explains (at least part of) the Bradley effect, but a post-hoc allocation of the undecideds would thus entirely miss a significant part of this phenomena. Moreover, in examining actual polling and elections, we do not have the luxury of knowing the ultimate vote choice of the "undecided" proportion of the electorate. This is partly why correctly identifying a Bradley Effect is a tricky task, even with the 'correct' method, but it becomes nearly impossible when methods are employed which are not robust in the absence of an allocation of the undecideds.

It is important to reiterate that the influence of the undecided proportion is solely a result of the normalizing divisors used in #2 and #9. When the magnitude of these divisors differs significantly from election to election, then this can cause what large fluctuations in #2 and #9, even if all other factors are equal. So, if we wished to study a single instance of polling inaccuracy in isolation, the choice between the three measures would likely not matter much, but if we wish to compare across cases, the preceding

example should cause one to think twice before employing #2 or #9. This implies that, in terms of this aspect of *Criteria 4*, #5 does better than #2 and #9.

Table 2B illustrates the effects of differences in the competitiveness of the race (measured by the predicted margin of victory) on the two measures. Table 2B displays three hypothetical elections with a one point difference between the major party vote share in the poll and in the election. For each of these cases, candidate A performs one point better in the actual result and candidate B loses a point from the poll to the election. What varies in this example are the predicted and actual margins of victory. In the first case, A loses by the largest margin, the second case is more competitive, and in third candidate A holds a large lead over candidate B. Regardless of the margin of victory, #5 shows that a one percent improvement for A and a one percent decline for B in the pre-election polls leads to a 2% "reverse" Bradley Effect (i.e., candidate A exceeded her predicted margin of victory by 2%).

Differences in the pre-election margin of victory can lead to large swings in the value of #2, even if there are few undecided voters and very little third-party support, as is the case in our example. Here, as the margin of victory for candidate A increases, #2 and #9 deviate from #5 more significantly. This is problematic, because using #2 or #9 in non-competitive states will produce a much larger Bradley Effect (or reverse Bradley Effect) than actually exists. In light of this, if one employs #2 or #9 in their analysis, meaningful comparisons cannot be made among elections varying in their competitiveness. These examples suggest that #2 and #5 do not fare well under yet another aspect of *Criteria 4*.

The normalizing divisors used in #2 and #9 are also meant to eliminate the effect of third-party support, in both the polling and voting data. The problem, however, is that unlike with the proportion of undecided voters—which can be 'eliminated' through some (arbitrary) exhaustive allocation to one or both of the candidates—there is no way to entirely eliminate the effect of the third-party related normalizing divisors, short of prohibiting third-party candidacies entirely. As a result, significant third-party support in either the poll or the election (or both) will also cause large fluctuations in the values of #2 and #9, while having no effect on #5.

In many Presidential (and other) races this point may be empirically moot as the sum total of support for the two major-party candidates is very near 100%. In some races, however, this is not the case. There are certain Presidential elections, such as 1992 with Ross Perot, in which there are serious third-party candidates. Even more problematic would be to apply #2 to Presidential primaries (especially in the early part of the primary season) because in these elections the share of the top two candidates is often *much* less than one. The difference in major party vote share influences #2 and #9 because it changes the relative magnitude of the denominator. With a smaller denominator, a change in polling inaccuracy of any given size is magnified.

Table 2C illustrates this point by showing how a one point discrepancy in the poll leads to large differences in #2 and #9 when the major parties' support decreases from 98% of the total electorate to 76% to 52%. As the major party vote share decreases, this one point change leads #2 and #9 to overestimate the degree of polling inaccuracy compared to #5. Notice that this problem is not eliminated in this case by an allocation of the undecideds. In fact, allocating the undecideds (at least under a 50-50 split) only serves

to exacerbate the effect of the change in third-party support. This example demonstrates that allocation of the undecideds—which is required to get a consistent measure for #2 and #9—can have profound effects on our estimate of the Bradley effect.

What this exposition makes clear is that, if we want to compare the accuracy of polling across many elections, to the extent that there are undecided voters in the poll or third-party support in the poll or in the election, we cannot rely on either #2 or #9 to give us measures appropriate for comparison. This is true even if an ex-post allocation of the undecided proportion is carried out. The only set of circumstances under which we can be sure that #2 and #9 will yield consistent, comparable figures is the unlikely one in which we have no third-party support or undecided voters.

Criterion 5: The Method is intuitive, readily interpretable, and is frequently used in academic and non-academic discussions of polling inaccuracy:

The final criteria we consider here is quite an important one, but it is also more subjective. What is a more "intuitive" measure of polling inaccuracy? Despite the nice properties of #9 as demonstrated by Martin, Traugott and Kennedy (2005), it seems unlikely that the most intuitive and readily interpretable measure of polling accuracy would involve a log-odds ratio. Moreover, #9 has not been, to the best of our knowledge, applied to the Bradley effect or a related phenomenon. Finally, #9, because it looks at the differences between the odds, for a particular candidate, of winning in the poll and the odds of winning in the election, it is implicitly invoking some notion of the 'winner' and thus is a counterintuitive measure for the application we are concerned with here.

There are numerous studies which use #5 (often along with #3, a non-candidate specific measure) as a method of measuring polling inaccuracy in general (see footnote 2

above). Moreover, Mitofsky (1998) notes that #5 is the statistic most often repeated by the media. #5 thus appears to be a more frequently used measure of general polling inaccuracy, but is this also the case for the specific application in which we are interested? In terms of the utility and intuitiveness of measures applied to the Bradley effect, the most well-known examples of the Bradley effect—such as Bradley's candidacy in 1982 and 1986 and Wilder's candidacy in 1989—are almost always implicitly discussed in terms of #5.

Since many discussions of polling inaccuracy in general, and especially those regarding a phenomenon such as the Bradley effect, explicitly or implicitly rely on the logic of margins of victory (which is the basis of #5), we believe that #5 also best meets *Criterion 5*.

Nonetheless, both #5 and #2 have been employed to measure the Bradley effect across a large number of heterogeneous elections (Hopkins, forthcoming, and Stromberg, 2008 employ #2; Stout and Kline, 2008, use #5). In the next section, we focus on the underlying relationship between #2 and #5, and subsequently investigate the real-world implications of the choice of method for measuring polling accuracy using polling and election data from all 50 states in the 2008 Presidential election.

Comparison of Measures: Data from the 2008 Presidential Race

To determine whether there was a Bradley Effect in the 2008 election and the differences in measuring the Bradley Effect between Mosteller #2 and #5, we collected pre-election polls and final election results for the presidential elections in all 50 states. We also collected the same data for states with Senate and gubernatorial elections. Pre-

election polls were collected from RealClearPolitics.com which archives polling data from several national and state polling firms. For each state, we used an average of all polls collected within the last week of the election (10/28/08-11/4/08). If this was unavailable, we used the most recent poll. For the election result, we collected data from each state's election authority.

Using both Mosteller #2 and #5 and the aforementioned data, we determine whether there was a Bradley Effect in the 2008 presidential election and whether these two methods differ on their assessment of this phenomenon. Although #9 is nearly as viable a measure as is #2 according to our criteria put forth above, we have omitted #9 from the following discussion for two reasons. First, we are not aware of #9 being employed in any studies of the Bradley effect or related phenomenon (thus it does not fare quite as well as #2 with respect to *Criterion 5*). Second, the behavior of #9, vis-à-vis #5, is qualitatively similar to that of #2. Thus, we discuss only the systematic divergence between #2 and #5 in what follows, but many of the claims would be similar if we were to discuss the differences between #9 and #5.

To make comparisons between methods we standardize both #2 and #5 scores for each state. Using these estimates, we first examine the average Bradley Effect for all 50 states and apply a two-sample difference of means test. Second, we disaggregate the state elections into sub-samples and re-test whether there were differences in states when you disaggregate by margin of victory, % undecided, and total party vote share.

The wide variation in the predicted margin of victory allows us to disaggregate into a bottom quartile (States were Obama was losing by 11.4% or more), an inter-quartile range and a top quartile (States were Obama was winning by 15.5 or more). For

percent of undecided voters, there is less variation: in fact the maximum number of undecided voters in any poll was only 9%. Due to the lack of differences between the number we separate the states by lowest 20 percent (% undecided range from 0-1.9), middle 60 percent (2 to 4%), and top 20 percent (% undecided ranges from 4.1 to 9). Vote share has the least variation; only 3 percent separated the state with the lowest major party vote share states with the highest. Therefore, we disaggregate into approximate halves (The lower range has a major party vote share of less than 99% and the higher range is greater than 99%).

[Insert Table 3 About Here]

As expected, neither #2 nor #5 indicate that there was a Bradley Effect in the 2008 presidential election. While there was a sizable Bradley Effect in states such as California, Arkansas, and New York, other states such as Vermont, Rhode Island, and New Mexico had polls that underestimated Obama's support. On average polls were very accurate and a vast majority of states predicted the final result within the margin of error. On average, #5 shows that polls underestimate Obama's support by about a quarter of a percent. #2 shows that polls were essentially correct. There were no significant differences between the two methods using the two sample t-tests.

[Insert Table 4 About Here]

The next step in the analysis is to investigate whether the differences observed in the hypothetical data sets above are consistent with the differences we find when analyzing data from the 2008 election. Figure 1A shows the absolute difference between standardized #2 and #5 scores for all 50 states graphed against the predicted margin of

victory. The graph along with table 2B confirms our findings from the previous section about the margin of victory and differences between #2 and #5. As the election becomes less competitive (in either direction), #2 underestimates polling inaccuracies for Obama (compared to #5).

[Insert Figure 1 About Here]

Figure 1B also shows the absolute difference between standardized #2 and #5, but it takes into account the proportion of undecided voters. The trajectory shows that as the proportion of undecideds increase in any state, #2 again underestimates the size of the Bradley Effect. While there are no significant differences in table 2C, the data shows that as the number of undecided grows, #2's measurements becomes increasing smaller than #5. Finally, the lack of difference between states with regards to major party vote share makes any meaningful analysis of the difference between # 2 and # 5 difficult. In every state, Obama and McCain's support combined totaled at least 97%. The results are indicative of this fact and there is no difference between #2 and #5 in either figure 1C or table 2D.

We should note that, given that the multiple 'elections' among which we were making comparisons, had much less variation than would a sample of say, African American candidates for Senate and Governor. Despite the little variation that we observed among our three variables of interest, we nonetheless observe differences between the measures which are consistent with what we would expect given our theoretical analysis of the measures as constructs for measuring polling error in the context of the Bradley effect.

Discussion

The 2008 Election was a watershed election for underrepresented groups in American politics. For the first time in American history, a woman and a black candidate were the frontrunners in the Democratic Primaries. In the general election, a black candidate represented the Democratic Party and a woman vice-presidential candidate represented the Republican Party. In addition to racial/ethnic minorities making advances, there were ballot measures across the United States that pertained to LBGT rights. This historic election has already inspired several research projects that focus on the polling for these candidates and propositions.

Undoubtedly, future studies will continue to assess the accuracy of polling support for individual candidates, parties, or issues. These studies may consist of measuring the accuracy of poll support for woman, African American, Latino, LGBT, or Asian American candidates. These studies may also examine the inaccuracies in polling support for issues pertaining to these groups, such as the gay marriage ban proposal in California where polls underestimated support for the ban. Outside of the United States, scholars may be interested in measure support for a particular party such as the Austrian Freedom Party and Le Pen's National Front in France.

Previous studies of polling inaccuracy focus on polling inaccuracy, and not candidate specific polling inaccuracy. In this paper, we compared the most commonly used measurement for polling error for a single candidate (#2 and #5), and found that #5 is superior for several reasons. We have demonstrated that #2 is not as accurate in assessing non-competitive elections and elections with a large proportion of undecided voters in both the hypothetical data and the 2008 election. Unfortunately, due to lack of variation in the 2008 election data we were only able to show how changes in the major

party vote share affect #2 with the hypothetical data. Signaling that the differences in major party vote share may not be a concern in general elections.

#2 overestimates the Bradley Effect in non-competitive elections. Therefore, large scale comparisons between candidates would be difficult. The same level of change from the polls to the election will not be treated equally with #2. Therefore, researchers using this method with minority candidates such as black Senate and gubernatorial candidates (who tend to be in non-competitive elections [Stout and Kline, 2008]) may have biased results. #5 is also superior for those analyzing primary data with minority candidates, especially when there is a comparison between only two candidates. In most primaries, especially in early states, the two most competitive candidates' support combined will be less than 100%. By only examining two candidates in elections where a third (not to mention fourth) candidate garners non-trivial support, #2 may be problematic.

Most importantly, even under ideal conditions (when there are only negligible proportions of undecided voters and third-party support), #2 does not give us a better measure, *in any single respect*, than #5. This is because, in this ideal scenario, #2 and #5 are equivalent up to a constant of proportionality and thus the value of one implies a value of the other. So, even if the actual differences between the two measures are small, as is the case for the 2008 data we analyzed above, there is not a single dimension along which #2 is superior to #5, though, as our preceding analysis shows, the converse of this claim is not true. If not for any other reason, #5 is superior to #2 because it allows us to compare the magnitude of the Bradley effect across multiple elections.

In addition to issues with consistency over cases, #5 is also a more practical way to convey polling inaccuracies to the public. When the media discusses polling biases

they often refer to the number of percentage points by which the poll over or underestimated support for a particular candidate, rather than to a change in their vote share. With an increase in public interest in Bradley Effect and similar phenomena, the use of #5 would make this research more accessible to non-academics.

Table 1: Nine Methods for Measuring Polling Inaccuracy

$\#1 \equiv P_A - V_A$, where A is the leading candidate or candidate of interest.

$\#2 \equiv \frac{P_A}{P_A + P_B} - \frac{V_A}{V_A + V_B}$, where A is the leading candidate, or the candidate of interest.

$\#3 \equiv \frac{\sum_{i=1}^{n} |P_i - V_i|}{n}$, where n is the number of candidates in the election.

$\#4 \equiv \frac{\sum_{i=1}^{n} \left|1 - P_i / V_i\right|}{n}$, where n is the number of candidates.

$\#5 \equiv (P_A - V_A) - (P_B - V_B)$, where A is the candidate of interest, or the leading candidate.

$\#6 \equiv \max_{i=1,2,\ldots n} |P_i - V_i|$, where n is the number of candidates/parties in the election.

$\#7 \equiv$ A χ^2 test of congruence, based on an $n \times n$ table, with P_i the row variable and V_i the column variable.

$\#8 \equiv$ "the difference between the predicted and actual electoral (college) vote. (Mitofsky, 1998 p.)"

$\#9 \equiv \log\left[\left(P_A / P_B\right) / \left(V_A / V_B\right)\right]$, or a log-odds ratio where A and B are the top two candidates.

Where P_i is the percentage of poll respondents indicating their intent to vote for candidate i, and V_i is the percentage of votes for candidate i, and, if we consider only two candidates, i = A, B. In our application, candidate A is our candidate of interest, i.e. the candidate with respect to whom we are assessing the Bradley effect.

Table 2: Comparison of Mosteller #2 and #5, Various Hypothetical Scenarios													W/o Undecided Allocation			W/Undecided Allocation		
	Pa1	Pb1	Pa1+ Pb1	Und	Pa2	Pb2	Poll1 Mov	Vote A	Vote B	Va+ Vb	3rd-Party %	Vote MOV	#9	#2	#5/2	#9	#2	#5/2
A	0.4	0.45	0.85	0.15	0.475	0.525	-0.05	0.425	0.575	1	0	-0.15	0.18	4.56	5	0.2	5.0	5.0
	0.25	0.6	0.85	0.15	0.325	0.675	-0.35	0.275	0.725	1	0	-0.45	0.09	1.91	5	0.2	5.0	5.0
B	0.14	0.83	0.97	0.03	0.155	0.845	-0.69	0.16	0.83	0.99	0.01	-0.67	-0.13	-1.73	-1	-0.05	-0.66	-1.0
	0.49	0.48	0.97	0.03	0.505	0.495	0.01	0.51	0.48	0.99	0.01	0.03	-0.04	-1.00	-1	-0.04	-1.02	-1.0
	0.83	0.14	0.97	0.03	0.845	0.155	0.69	0.85	0.14	0.99	0.01	0.71	-0.02	-0.29	-1	-0.11	-1.36	-1.0
C	0.23	0.74	0.97	0.02	0.24	0.75	-0.51	0.24	0.74	0.98	0.02	-0.5	-0.04	-0.78	-0.5	-0.01	-0.25	-0.5
	0.12	0.63	0.75	0.02	0.13	0.64	-0.51	0.13	0.63	0.76	0.24	-0.5	-0.08	-1.11	-0.5	-0.02	-0.22	-0.5
	0	0.51	0.51	0.02	0.01	0.52	-0.51	0.01	0.51	0.52	0.48	-0.5	..	-1.92	-0.5	-0.02	-0.04	-0.5

Table 3) Comparison of Means Test using the 2008 Presidential Election Data Across 50 States

3A) Mosteller #2 & #5 assessment of 2008 Presidential Election across 50 States

	Obs	Mosteller 2	Mosteller 5
Average	50	-0.01%	-0.24%

3B) Mosteller #2 & #5 assessment of 2008 Presidential Election and Predicted Margin of Victory

	N	Standardized Mosteller 5 Average	Standardized Mosteller 2 Average
Bottom Quartile	12	-.27*	-0.11*
Middle Half	26	-0.013	-0.011
Top Quartile	12	0.29*	0.1*

3C) Mosteller #2 & #5 assessment of 2008 Presidential Election and % Undecided

	N	Standardized Mosteller 5 Average	Standardized Mosteller 2 Average
Bottom 20%	10	0.49	0.49
Middle 60%	30	0.65	0.63
Top 20%	10	0.86	0.76

3D) Mosteller #2 & #5 assessment of 2008 Presidential Election and Major Party Vote Share

	N	Standardized Mosteller 5 Average	Standardized Mosteller 2 Average
Approximately Bottom Half	18	0.79	0.78
Approximately Top Half	36	0.55	0.53

*Significant at .05

Table 4: Comparison of Means Test measuring the Bradley Effect for Barack Obama vs. the 2008 Democratic Gubernatorial and US Senate Candidates

Candidate	Obs.	Mosteller 5	Mosteller 2
Obama	35	-0.15	-0.06
White Democratic Comparisons	35	-0.81	-0.16
Difference		0.66	0.1

* Significant at .05

Figure 1A) Predicting the Difference in the Absolute Standardized Values of #5- #2 vs. Pre-Election Margin of Victory

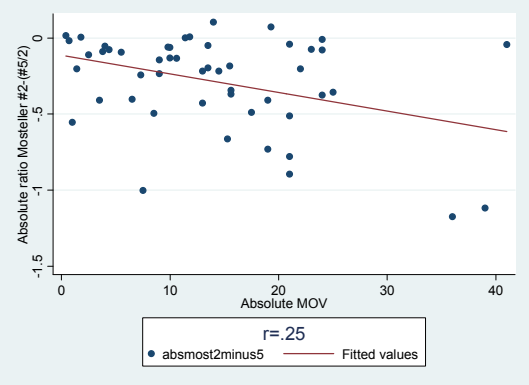


Figure 1B) Predicting the Difference in the Absolute Standardized Values of #5- #2 vs. % Undecided

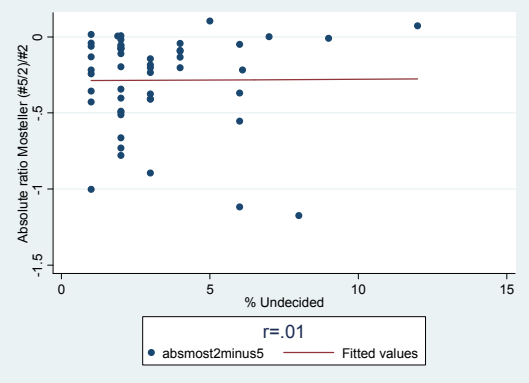
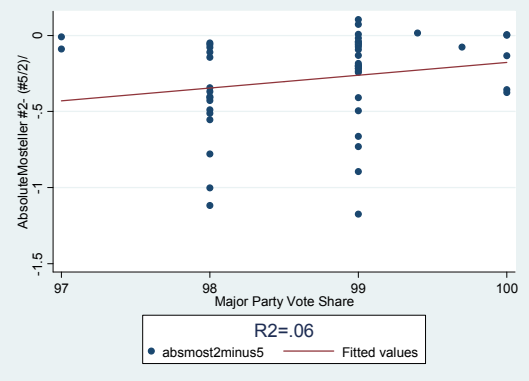


Figure 1C) Predicting the Difference in the Absolute Standardized Values of #5- #2 vs. Major Party Vote Share



References

Berinsky, Adam J. (1999). "The Two Faces of Public Opinion." American Journal of Political Science 43(4): 1209-30.

Crespi, Irving. (1988). Polling Sources of Accuracy and Error, New-York: Russell Sage.

Hopkins, Daniel (forthcoming). "No More Wilder Effect, Never a Whitman Effect: When and Why Polls Mislead about Black and Female Candidates." Journal of Politics.

Keeter, Scott, & Samaranayake, Nilanthi. (2007). "Can you trust what polls say about Obama's electoral prospects?" Retrieved March 12, 2008, from http://pewresearch.org/ pubs/ 408/can-you-trustwhat-polls-say-about-obamas-electoral-prospects.

Malghaes, Pedro C. (2005). "Pre-Election Polls in Portugal: Accuracy, Bias and Sources of Error, 1991-2004." International Journal of Public Opinion Research. 17(4): 399-421.

Martin, Elizabeth A., Michael W. Traugott and Courtney Kennedy. (2005). "A Review and Proposal for a New Measure of Poll Accuracy." Public Opinion Quarterly. 69(3): 342-369.

Mitofsky, Warren. J. (1998). "Review: Was 1996 a Worse Year for Polls Than 1948?" Public Opinion Quarterly 62(2): 230-249.

Mosteller, Frederick. 1949. "Measuring the Error." In *The Pre-election Polls of 1948*, by Frederick Mosteller et al. New York: Social Science Research Council.

Panagakis, Nick. (1999). "Response to 'Was 1996 a Worse Year for Polls Than 1948?' Public Opinion Quarterly 63(2): 278-81.

Stout, Christopher and Reuben Kline. (2008). " Ashamed Not to Vote for an African-American; Ashamed to Vote for a Woman: An Analysis of the Bradley Effect from 1982-2006." Center for the Study of Democracy Working Paper. 08-08.

Stromberg, David. (2008). "How Large is the Bradley Effect and Does it Matter for Obama?" Vox. http://www.voxeu.org/index.php?q=node/2514. November 3, 2008.

Traugott, Michael W. (2001). "Trends: Assessing Poll Performance in the 2000 Campaign." Public Opinion Quarterly 65(3): 389-419.

Traugott, Michael W. (2005). "The Accuracy of the National Preelection Polls in the 2004 Presidential Election." Public Opinion Quarterly 69(5): 642-654.

Traugott, Michael W. and Vincent Price (1992). "A Review: Exit Polls in the 1989 Virginia Gubernatorial Race: Where Did They Go Wrong?" The Public Opinion Quarterly 56(2): 245-253.

Appendix 1:

Need to Insert Proof that #5=2*#2 iff Pa+Pb=Va+Vb=1