

No More Wilder Effect, Never a Whitman Effect: When and Why Polls Mislead about Black and Female Candidates

Daniel J. Hopkins*
Post-Doctoral Fellow
Department of Government
Harvard University
dhopkins -at- iq -dot- harvard -dot- edu

August 4, 2008

*This research was supported by Yale University's Center for the Study of American Politics. The author is especially grateful to Elizabeth Campbell for research assistance, to Samir Jenkins for sharing data on the 2008 primaries, and to Stephen Ansolabehere, Adam Berinsky, Barry Burden, Alan Gerber, Emily Gregory, Tony Greenwald, Justin Grimmer, Gregory Huber, Gary King, Daniel Schlozman, and John Sides for advice, assistance, or suggestions. This work was presented as a poster at the 2008 Meeting of the Society for Political Methodology.

Abstract

The 2008 election has renewed interest in the Wilder effect, the gap between the share of survey respondents expressing support for a candidate and the candidate's vote share. Using new data from 133 gubernatorial and Senate elections from 1989 to 2006, this paper presents the first large-sample test of the Wilder effect. It demonstrates a significant Wilder effect only through the early 1990s, when Wilder himself was Governor of Virginia. Although the same mechanisms could affect female candidates, this paper finds no such effect at any point in time. It also shows how polls' over-estimation of front-runners' support can exaggerate estimates of the Wilder effect. Together, these results accord with theories emphasizing how short-term changes in the political context influence the role of race in statewide elections. The Wilder effect is the product of racial attitudes in specific political contexts, not a more general response to under-represented groups.

From the vantagepoint of 1989, the existence of a “Wilder effect” or “Bradley effect”—a gap between how black candidates polled and how they performed—was hard to dispute. That year, Virginia gubernatorial candidate Douglas Wilder enjoyed a 15 percentage-point lead two weeks before the election (Melton and Morin, 1989), and yet won by only 6,700 votes out of more than 1.7 million cast. On the same day, New York’s David Dinkins won the Mayoralty by 2 percentage points after polls showed him leading by as many as 18 percentage points (Keeter and Samaranayake, 2007). After Tom Bradley’s surprising loss in the 1982 California Governor’s race, observers knew how to interpret these gaps: as evidence of subtle anti-black biases that voters were not willing to voice in polls but that were operative in the voting booth. In the words of the campaign manager for Bradley’s opponent, “[i]f people are going to vote that way, they certainly are not going to announce it for a survey taker” (Reeves, 1997, pg. 15).

Discussions of the Wilder effect¹ have been most common among journalists, but the effect is closely related to ongoing political science debates about the influence of racial attitudes on voters’ decisions. Measuring the Wilder effect over time will help us understand the roots, stability, and potency of one subtle form of anti-black electoral behavior. If there is a detectable Wilder effect, is it a durable phenomenon or one that appears only under certain political conditions? This paper will also consider—seemingly for the first time—whether such gaps between polling and performance affect female candidates. When Christine Whitman won the Governorship of New Jersey, her first public comments attacked polls that had *underestimated* her support (Jackson, 1993). That prompts a question: are Wilder effects a specific response to black candidates or a common response to candidates from under-represented groups? These questions take on special importance in 2008, a year which witnessed the most serious black and female contenders for a major-party Presidential nomination in U.S. history.

A handful of high-profile elections with African American contenders have dominated thinking about the Wilder effect, including those involving Wilder (1989), Bradley (1982), and Dinkins (1989, 1993) (Keeter and Samaranayake, 2007; Bishop and Fisher, 1991; Traugott and

¹This paper refers to the “Wilder effect” for consistency, but considers the “Bradley effect” an exact synonym.

Price, 1991). But given that polls are imprecise and administered prior to the end of the campaign, one could appeal to idiosyncratic explanations in any given case. Since political scientists appear not to have studied similar effects in large samples of elections or for other under-represented groups, we know little about the polling-performance gap and its typical properties. Perhaps wide swings in the final weeks of an election cycle are common, and the handful of well-known examples of the Wilder effect are counter-balanced by less visible cases where black candidates gained substantially on their polling numbers. Polls overstate support for leading candidates (Erikson and Wlezien, 2007; Wlezien and Erikson, 2002), making it possible that subsequent outcomes will exhibit patterns similar to regression to the mean. Or perhaps the gap between polling and performance genuinely represents a subtle anti-black bias.

To assess these competing explanations, using a broad set of observations is crucial. This paper compiles polling and election data for all black and female candidates for Governor or U.S. Senator from 1989 to 2006. These 249 observations from 133 elections show that *there was indeed a Wilder effect, but one that was specific to a particular group and political context*. African Americans running for office before 1996 performed on average 2.7 percentage points worse than their polling numbers would indicate. Yet this effect subsequently disappeared. Although precision is limited because there were only 47 observations for 18 elections with black candidates in this period, these findings accord with theories of racial politics emphasizing the information environment. As racialized rhetoric about welfare and crime receded from national prominence in the mid-1990s, so did the gap between polling and performance. Even over short periods of time, the influence of race on electoral politics can shift markedly. Moreover, when it was influential, the Wilder effect was specific to black candidates. To borrow the name of the former New Jersey Governor, we observe a Wilder effect but never a Whitman effect.

Section 1 defines the Wilder effect and outlines potential explanations for it. That section also contends that estimating the Wilder effect can help assess theoretical arguments about how racial attitudes shape voting behavior. In Section 2, the paper discusses data selection and compilation before estimating the Wilder effect for black and female candidates in Section

3. The Wilder effect has not been scrutinized thoroughly in the recent past, so this section familiarizes readers with these types of data. The tests presented here reaffirm that polls typically overstate support for front-runners. Irrespective of their race, when candidates have high levels of initial support, we should expect their performance to decline come election time. This can lead naive estimates of the Wilder effect to overstate its magnitude. A final test shows no Wilder effect in the 2008 Democratic presidential primary.

1 Theory

The Wilder effect is commonly defined as the difference between the share of the electorate voicing support for a black candidate in a survey and the share casting ballots for that candidate. In certain elections, it has been as large as 16 percentage points (Keeter and Samaranayake, 2007). A variety of factors could produce such a gap. One is last-minute campaign events, which might change the electorate’s preferences after the final surveys. Another is unexpected voter turnout or biased survey sampling, either of which could lead the electorate and the surveyed population to diverge. As Hajnal (2007, pg. 43) has demonstrated, turnout is often high when black challengers compete with white candidates. Still, the most common interpretation of the gap focuses on the interplay of racial bias and social desirability (e.g. Keeter and Samaranayake, 2007; Reeves, 1997). During a survey, respondents may be less likely to indicate voting intentions that are socially stigmatized, such as their unwillingness to support a black candidate (Berinsky, 2004, 1999; Krysan, 1998; Reeves, 1997). Put differently, the Wilder effect is a subset of the mode effects commonly discussed by survey researchers: the answers people give depend on how they are asked (e.g. Fowler, Roman and Di, 1998; Aquilino, 1994; Traugott and Price, 1991) and by whom they are asked (e.g. Bishop and Fisher, 1991; Finkel, Guterbock and Borg, 1991). Different modes impose different social pressures.

This paper defines the Wilder effect narrowly, excluding turnout effects and campaign effects. To the extent possible, it focuses on the gap stemming from respondents’ unwillingness to give socially undesirable answers in phone surveys immediately prior to the election.

Traugott and Price (1991) provide survey-based evidence for this effect, demonstrating that whites in the 1989 Virginia election were more likely to report voting for Wilder when asked in face-to-face exit polls. Here, the goal is instead to quantify these effects for as many separate pre-election polls as possible, and to test hypotheses about variation in the Wilder effect.

To be sure, there are many ways that stereotypes about a group could influence voters. The Wilder effect might not even be operative given the number of respondents who freely admit their prejudices to researchers (Citrin, Green and Sears, 1990) or who vote irrespective of the candidate's race (Highton, 2004). But the Wilder effect is of particular interest because it influences election outcomes in unanticipated ways, and because it fits closely with theories of egalitarian norms and symbolic racism. According to these arguments, voicing explicitly racist views has become socially unacceptable since the 1950s (Mendelberg, 2001; Sears et al., 2000; Sears, Henry and Kosterman, 2000; Kinder and Sears, 1981). Racism's influence has taken on subtle forms, as whites use traditional, race-neutral values to justify discriminatory actions (Frey and Gaertner, 1986) and express anti-black attitudes (Sears, Henry and Kosterman, 2000). Here again, we see voters expressing themselves differently as social pressures change. The Wilder effect is one observable implication of theories of subtle or symbolic racism.

Not all studies agree with the idea of symbolic racism, or its conflation of racial attitudes and conservative political ideology (e.g. Sniderman, Crosby and Howell, 2000; Sniderman and Carmines, 1997). Still, research on both sides of the debate has drawn primarily on cross-sectional survey data and survey experiments, leaving us unsure of how subtle forms of bias vary over time. As a measure that is comparable across elections, the Wilder effect is potentially valuable in understanding to the strength and persistence of symbolic or subtle racism. Is there really a Wilder effect? If so, has the Wilder effect varied with time? These are at once questions about surveys' accuracy for black candidates and questions about symbolic racism.

Older whites are more likely to hold negative views of blacks (Schuman et al., 1997), so the gradualism hypothesis holds that Wilder effects will decline gradually as younger voters replace older voters. Another hypothesis holds that racist responses are more common when

white voters lack positive information about black candidates, and allow stereotypes and uncertainty to fill the void (Hajnal, 2007). Thus the strength of the Wilder effect might hinge on the information environment, and might decline when the black candidate is a familiar incumbent or when other available information diminishes whites' fears. Other scholarship has also pointed to the role of the information environment, albeit in a different way, demonstrating how racialized campaigns prime anti-black attitudes (e.g. White, 2007; Huber and Lapinski, 2006; Mendelberg, 2001; Valentino, 1999; Reeves, 1997; Citrin, Green and Sears, 1990). A focus on racially charged issues can increase the importance of racial considerations in voting as surely as a focus on incumbency or performance can decrease their importance. To the extent that racial considerations disproportionately influence evaluations of black candidates, the information environment could be a critical moderator of the Wilder effect.

During a racially charged campaign, a candidate's race will be foremost in voters' minds. According to this racialization hypothesis, we might expect abrupt changes in the Wilder effect as the information environment shifts. The focus in this paper is on the *national* information environment, as it has observable implications for many elections, although campaigns certainly can racialize local and statewide streams of information as well. In the period under study, the prominence of racialized issues such as crime and welfare declined markedly at the national level in the late 1990s and early 2000s (Jones and Baumgartner, 2005). In 1995, 12% of Americans cited social welfare issues as the nation's most important problem, a figure that was just 5% by 2001 and 4% in 2004 (Baumgartner et al., 2006). In 1994, 29% of Americans cited crime as the nation's most important problem, a figure that had dropped to 9% by 2004. Both of these issues are closely connected to racial attitudes (e.g. Mendelberg, 2001; Gilens, 1999; Valentino, 1999; Hurwitz and Peffley, 1997), yielding a prediction: for black candidates, the Wilder effect should have declined during this period. In the early 1990s, white voters would have been primed by national politics to connect blacks and black candidates to issues such as crime and welfare. By the turn of the 21st century, those associations were less salient nationally.

1.1 Alternative Explanations and Groups

There are other explanations for why we might observe a Wilder effect. In every four-year period, the U.S. elects more than 100 Senators and Governors. Assuming that the media pay particular attention to promising candidates from under-represented groups (e.g. Reeves, 1997), candidates come to public attention precisely because of high poll numbers. Yet in part, this selection process also means that candidates who come to public attention are more likely to be outperforming expectations—and are primed for a decline. Put differently, one might explain the observed instances of the Wilder effect as examples of *regression to the mean*. Alternately, to the extent that polls measure name recognition rather than likely voting behavior, they may systematically overstate the vote share of the better-known candidate, even if they are unbiased estimators of candidates' expected vote share on average. This is not classical regression to the mean, but a substantive process that produces a very similar result. Both processes lead to the expectation that candidates with polls that are especially high or low should expect election outcomes that are less extreme. Douglas Wilder might have seen his numbers decline because he was ahead, not because he was black.

The possibility of front-runner decline makes it valuable to examine non-black candidates, as there is nothing specific to black candidates in this hypothesis. There are other reasons to study multiple groups as well: by examining other groups subject to systematic biases, we can better understand whether the Wilder effect is specific to African Americans or general to under-represented groups. Clearly, female candidates face gender-based stereotypes and biases that respondents do not always admit to (Streb et al., 2008; Sanbonmatsu, 2002; Koch, 2002; Jamieson, 1995; Sapiro, 1981). As with black candidates, these biases are thought to be more pronounced in low-information environments or when certain frames are salient in the news media (Kahn, 1996; Jamieson, 1995; Alexander and Andersen, 1993), providing the potential for variation across states and years. Yet issues surrounding female candidates do not appear to provoke concerns about social desirability (e.g. Sanbonmatsu, 2002, pgs. 23,31) to the same extent as issues surrounding black candidates (Mendelberg, 2001), allowing us to observe if

stereotypes in the absence of strong norms can still produce a Wilder-type effect. Studying women provides an instrumental advantage as well. There are many more female candidates than Latino candidates or Asian American candidates, enabling us to measure phenomena such as front-runner fall-off with precision.

In the U.S., female elected officials are more commonly found in legislative as opposed to executive positions, perhaps because those roles accord with gender-based stereotypes about the issues on which candidates are most effective (Kahn, 1994; Alexander and Andersen, 1993; Huddy and Terkildsen, 1993; Rosenwasser et al., 1987). Research also shows that gender may interact with partisan stereotypes, as females are typically associated with liberal political positions (Koch, 2002). Given that, analyses of gender should differentiate between women seeking legislative posts and those seeking executive posts. They should also be attentive to whether there is a polling-performance gap only for members of one political party or any one point in time. We now turn to the data.

2 Data and Methods

To test whether the Wilder effect holds in large samples, whether it varies over time, and whether it affects female candidates, we need data on polls and general election outcomes for races involving black and female candidates. We also need data on potential confounding factors, including the office being sought, the candidate’s party, the poll’s date and sample size, whether the candidate was an incumbent, whether the opponent was an incumbent, the state’s average partisanship, the level of voter turnout, the change in voter turnout, and the state’s percent black. We obtained election returns and incumbency information from the online “Atlas of Presidential Elections” and the Clerk of the U.S. House of Representatives.² Demographic information came from the U.S. census.³ Turnout data came from Michael McDonald’s United States Election Project.⁴ The analysis here is restricted to Senate and

²Available at: <http://uselectionatlas.org> and <http://clerk.house.gov/> respectively.

³Available online at: <http://www.census.gov>

⁴Available online at: http://elections.gmu.edu/voter_turnout.htm.

Gubernatorial races from 1989 to 2006 due to the limited data availability for other offices and earlier years. To avoid the complexities of multi-candidate races, the analysis includes only candidates nominated by major parties.

We restricted the analysis to polls conducted by outside firms, excluding candidates' internal polling and polling done by political parties. There is no centralized compilation of this kind of polling data, so we compiled archived newspaper articles. Specifically, we searched national papers or the relevant state's papers using Lexis Nexis for the candidate's last name and the words "poll," "margin," or "survey." For each candidate in a given year, the analysis included the three most recent polls where multiple polls were available, although three or more separate polls were found in only 25% of cases. To be included, all polls had to be completed within one month of election day to limit fluctuations in support induced by campaign events or public inattention.⁵ In a handful of cases, the poll's completion date was not available, and was set to the date of the article's publication. Many of the newspaper stories do not indicate whether the survey sampled all adults, registered voters, or likely voters, and the screens for likely voters differ across survey organizations, so this information was not used.⁶

The analysis excludes the handful of campaigns where an African American or woman ran against someone from the same group. In such circumstances, a Wilder effect would be impossible to identify, as the average gap between polling and performance is by definition zero.⁷ For the remaining 151 black or female candidates during this period, we found at least one pre-election poll for 88%. All but one of the 18 missing candidates was female, and the vast majority were token challengers with few resources.⁸ In total, we found 208 polls for 119 female candidates and 47 polls for 18 black candidates. The polling metric is the share of respondents supporting a major-party candidate who support the female or black candidate.

⁵In the case of Louisiana run-off elections and a 1993 special election in Texas, the date was set to 30 days prior to the election.

⁶The polls were conducted by 141 different polling or news organizations, so it is impossible to estimate whether a particular polling firm is more or less accurate for black and female candidates. For more on estimating house effects, see Beck, Jackman and Rosenthal (2006).

⁷Among others, this criterion removes Barack Obama's 2006 Illinois Senate race against Alan Keyes because both are black and Susan Collins' 2002 Senate race against Chellie Pingree because both are women.

⁸The one African American without available polling data is Troy Brown Sr., who took 32% of the vote in his 2000 race against Mississippi incumbent Senator Trent Lott.

We constructed the election outcome metric in the same way, ignoring third-party candidates. Respondents who tell pollsters they are undecided and then disproportionately vote against the minority candidate (e.g. Berinsky, 1999) will increase the polling-performance gap.

[Table 1 here]

Tables 1 and 2 provide descriptive statistics for key variables. The “State Share Democratic” refers to the share of the national two-party vote received by the Democratic Presidential candidate in the most recent election above or below the national Democratic vote share. This measure of statewide partisanship removes national swings in the popular vote. For instance, the maximum of 0.14 among black candidates indicates that Massachusetts in 2004 favored the Democratic candidate by 14 percentage points more than did the nation as a whole. Turnout measures the share of the voting-eligible population who voted in that election (McDonald and Popkin, 2002), and the change in turnout is as compared to the election four years prior. Biracial elections do increase turnout as compared to the previous election by an average of 3 percentage points. These descriptive statistics also provide the first hints that the Wilder effect is small in size. The average gap between a black candidate’s share in a poll and her subsequent share of the vote is just 1 percentage point. For female candidates, the figure is -0.4 of a percentage point. Most of the observations are for Democratic candidates, and the data set for black candidates has very few observations of incumbents (3 of 47).

[Table 2 here]

3 Results

How to explain the absence of the Wilder effect in these data when elections in the late 1980s made its existence so indisputable? Figure 1, which plots the polling-performance gap by year alongside loess smoothing lines, provides a preliminary answer. In the first several years, all of the observed gaps are positive, meaning that black candidates consistently polled better than

they performed. The Wilder effect was at work. Yet in 1994, we see black candidates such as Washington’s Ron Sims performing worse than their polls for the first time. And as of the late 1990s, the Wilder effect disappears entirely. Even Tennessee’s 2006 Democratic nominee for Senate, Harold Ford Jr., experienced no Wilder effect after a negative television advertisement targeting him cued anxieties about inter-racial sex.⁹ Just prior to the election, a *USA Today* survey put Ford’s support at 48.4 percentage points, 0.2 percentage points lower than he would perform on election day. This weighs against hypotheses emphasizing the local information environment exclusively. In addition, the speed of the Wilder effect’s disappearance is evidence against the gradualism hypothesis. For female candidates, Figure 1 shows no consistent gap between polling and performance at any point in time. This holds for female candidates seeking both legislative and executive positions. In fact, the point estimate for female Senate candidates is usually negative, indicating that they slightly *outperform* polls.

[Figure 1 here]

We now interrogate the key finding—a Wilder effect for black candidates until the mid-1990s—more systematically. As the year when welfare reformed passed, 1996 marked a decline in the salience of racialized national rhetoric. It also represented a turning point for the Wilder effect. Before 1996, the median gap for black candidates was 3.1 percentage points, while for subsequent years it was -0.3 percentage points. Even for this relatively small sample, the p-value on the t-test that these two distributions have the same mean is 0.01.¹⁰ That rejects the possibility that the differences across years stem from chance alone. To overcome concerns about these results being driven by outliers, an additional test randomly removed four of the elections from the data on black candidates at each iteration and then estimated the difference between the two periods. The Wilder effect was larger in the pre-1996 period in 99.9% of such simulations. For black candidates, the early 1990s saw a pronounced Wilder effect—but

⁹Specifically, a white actress in the late-October advertisement exclaims “I met Harold at the Playboy party” and then closes the advertisement by winking and saying, “Harold, call me.”

¹⁰Even removing all but the most recent survey for each of the 18 elections, the p-value from this t-test remains 0.03. Thus the results are not an artifact of including polls too far out from the election or from clustering multiple observations within an election.

one that has vanished since. Parallel analyses confirm that female candidates see no such change over time, and that their polling-performance gap is consistently near zero or even negative. Indeed, female Senate candidates outperform their poll numbers by an average of 1.2 percentage points.¹¹

These patterns seem to confirm that the national information environment influences the strength of the Wilder effect. Still, several threats to validity remain, including omitted variable bias. One can easily imagine that the 18 observations of black candidates prior to 1996 differed in some important respect from the 29 subsequent observations, perhaps because the earlier candidates were more likely to be Democrats or running in more heavily black states (Greenwald, 2008). Also, we should not expect the Wilder effect to be constant for candidates with very different levels of baseline support. There are multiple reasons for this. Since candidates cannot lose support they never had, the Wilder effect might be an increasing function of a black candidate’s initial level of support. Alternately, regression to the mean or similar processes might be at work. Finally, the results above treat polling outcomes as fixed, ignoring sampling variability.

To address these issues, we began with a basic statistical model where the polling-performance gap Y_i for each observation i is a linear function of the number of days until the election, the candidate’s level of pre-election support, an indicator variable denoting whether the election was prior to 1996, and a normally distributed error term. Conditioning on the number of the days until the election ensures that these results are not driven by the polls furthest from election day, and thus by campaign events. We include pre-election polling to account for the possibility that the Wilder effect varies with the candidate’s initial support—and discuss this in more detail in Section 3.1 below.

Utilizing polling results as an independent variable introduces another challenge. The polls measure the distribution of public opinion with sampling-induced variability. We thus calculated the sampling variance of each poll given its result \hat{p}_i and sample size n_i :

¹¹A t-test confirms that this figure is different from zero at $p=0.01$.

$$\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i - 1},$$

as outlined by Rice (1995, pg. 198). For each observation, we then drew five simulations of the polling result \tilde{p}_i from a normal distribution with mean \hat{p}_i and the poll-specific variance. Following procedures for multiple imputation (King et al., 2001; Schafer, 1997), we estimated each model for the five resulting data sets. This procedure models sampling variability explicitly. At the same time, group-clustered standard errors account for the dependence of the observations within each election (Wooldridge, 2003).¹²

[Figure 2 here]

To check the robustness of the central finding, we added one variable at a time to the basic model and then extracted the coefficients for the time period effect and the new variable. This allows us to observe the influence of each new variable in models that are not overly saturated (see Achen, 2005). Figure 2 presents the results. Each dot represents the estimated coefficient, with the interval within one standard error given by a solid line and the 95% confidence interval given by a dashed line. Coefficients on the left represent the increased Wilder effect in the early 1990s when accounting for new variable whose coefficient is presented on the right. For example, the line at the bottom of the left graph indicates that when conditioning on whether the candidate is a Democrat, black candidates running before 1996 saw a polling-performance gap 2.6 percentage points larger than black candidates running in subsequent years. The corresponding line on the right indicates that the effect of being a Democrat is -1.8 percentage points in the same model, with a standard error of 1.0 percent point. Black Democratic candidates perform better given their polling numbers than black Republican candidates.

Figure 2 shows that the key finding is robust to a range of potentially omitted variables. Across several specifications, elections with black candidates before 1996 always generated a polling-performance gap at least 1.7 percentage points larger than elections after. This period

¹²There are only one to three observations per election, and results from multilevel models are not reported because they did not always converge.

effect is not significantly smaller when accounting for the sample size of the poll, whether the election saw increased turnout, whether the election saw high turnout, whether the state leans Democratic, the state's percent black, whether the election was in the South, whether the opponent was an incumbent, the position sought, or the candidate's party. Importantly, voter turnout itself is not a strong predictor of the polling-performance gap. Also, including the change in voter turnout from four years prior increases the estimated period effect, to 3.3 percentage points. Together, these two findings weigh decisively against the claim that the earlier polls were inaccurate due to unexpectedly high turnout in black-white elections.¹³ States with a higher share of African Americans tend to see larger polling-performance gaps, a finding that differs from what Greenwald (2008) observes in the 2008 Democratic primaries. Still, the central conclusion proves quite robust. Black candidates running in the early 1990s could expect a marked decline from their polling numbers to their actual performance.

3.1 Overstating Front-runner Support

Exploratory analyses uncovered an aspect of these data that proves critical in isolating the Wilder effect. The main candidates who suffered from the Wilder effect were all front-runners, including Bradley, Dinkins, and Wilder. If front-runners commonly experience a decline from their surveys to their performance at the ballot box, the racial aspect of the Wilder effect might be overstated. To check this possibility, the analysis first turned to female candidates, a group that does not suffer from any polling-performance gap. Among candidates polling above 50%, the average polling-performance gap was 1.9 percentage points, meaning that they tended to do worse at the polls than surveys would predict. For female candidates who were behind in the polls, they performed an average of 2.5 percentage points better on election day. The polling-performance gap depends on a candidate's initial level of support.

[Figure 3 here]

¹³Comparable voter turnout figures are not available for the 7 observations that occurred in off-year elections, including those in Louisiana in 1995 and 1999. Sample size information was missing for 10 polls. For these variables, missing data were multiply imputed.

The correlation between female candidates’ polled level of support and the polling-performance gap is an impressively strong 0.63. Very few female candidates with low poll numbers see their vote share decline, and very few female candidates with high poll numbers see their vote shares increase. The left half of Figure 3 depicts the results graphically. Drawing on a linear model, it shows the polling-performance gap as a function of the candidate’s support in the survey for both 1990-1995 and 1996-2006. The lines for the two periods are indistinguishable, reflecting the absence of a period effect among female candidates. And the pattern of growing gaps as initial support increases holds in both periods. The right half of Figure 3 shows the same pattern for black candidates, with a seemingly identical slope. There, too, the size of the Wilder effect hinges on a candidate’s initial level of support. The gap between the two lines illustrates the difference between the polling-performance gap before 1996 and after, reinforcing that the gap for black candidates was much wider prior to 1996.

What might cause such patterns? Earlier, we suggested that the Wilder effect might grow with a black candidate’s initial level of support. But if that were the case, the polling-performance gap should be nearly zero for a candidate with little support. Yet here it is decidedly negative for such candidates. Another possibility is classical regression to the mean, defined as “the phenomenon that a variable that is extreme on its first measurement will tend to be closer to the center of the distribution for a later measurement” (Davis, 1976, pg. 493). For classical regression to the mean to operate, the two measurements must be of random variables with the same mean and some non-zero variability. Under these conditions, if one observes a selected sample—for instance, including all those who score above average—we should expect scores to decline on average in the second measurement. Since polls have random variability induced by sampling, this is a plausible mechanism. However, simulations show that this is not the main source of front-runners’ observed decline.¹⁴

¹⁴Specifically, assume an unbiased poll that differs from the actual election outcome only due to sampling variability. Using the sample sizes for female candidates to generate random deviations around the election outcomes, we find that candidates performing above 50% in the simulated poll will regress 0.5 of a percentage point on average. This simulation isolates the fraction of the regression that is due to sampling variation. Since in actuality women performing above 50% in polls see an average decline of 1.9 percentage points, sampling variability accounts for roughly 25% of the observed front-runner decline.

Alternately, the pattern may have a substantive rather than statistical explanation. When faced with an unknown candidate, people in surveys may be more likely to voice support for a familiar front-runner, and might not anticipate the influence of partisan cues while in the voting booth. For our purposes, it is less important to explain this phenomenon than to account for it in estimating the Wilder effect. We do not want to attribute to anti-black bias what is actually an artifact of survey measurement. To estimate the expected decline for a front-runner, we again modeled the polling-performance gap as a linear function of the candidate's pre-election support and other covariates of interest. We use the initial model specification above, and add the candidate's party and the state's percent black given their strong predictive power in Figure 2. We can then use the model to predict the polling-performance gap for a given level of initial support. Figure 4 presents the results, highlighting Wilder's 1989 race for Governor of Virginia. The naive polling-performance gap in that election was 8.2 percentage points. Decomposing that effect, we know from the linear model that we should attribute 2.3 percentage points to the period effect of running in the late 1980s, when Willie Horton was a recent memory and when racialized rhetoric around crime and welfare were salient.

[Figure 4 here]

To calculate how much of Wilder's election-day decline was due to his status as a front-runner, the analysis used the linear model to predict the polling-performance gap at Wilder's observed level of support (58%) and also at 50%, indicating two candidates who are exactly tied. By definition, there is no front-runner effect for a two-candidate race with each polling at 50%. We set other variables to their medians. For a black candidate polling at 58%, the model predicts a polling-performance gap of 3.2 percentage points excluding the period effect. By contrast, had the same candidate been polling at 50%, the drop-off on election day would have been just 0.9 percentage points. Thus we can reasonably attribute 2.3 percentage points of the polling-performance gap to the standard front-runner decline. Re-estimating the same predictions for female candidates generates exactly the same correction of 2.3 percentage points, indicating that this level of front-runner decline is not specific to black candidates.

There was certainly a Wilder effect in Virginia in 1989, but in all likelihood, there was also the usual decline in front-runner performance.

3.2 A Final Test: 2008

The 2008 Democratic presidential primaries renewed speculation about the Wilder effect, so as a final test, we applied the same decision rules as above to collect up to three polls for each of the 33 U.S. states that held contested Democratic primaries.¹⁵ Doing so yields 87 observations of polled and actual support for Senators Barack Obama and Hillary Clinton. The mean polling-performance gap was 1.4 percentage points for Senator Clinton, and the reverse for Senator Obama. This estimated mean is not at all sensitive to particular polls or states: if we remove the observations for five randomly chosen states at a time, we still observe that Senator Obama’s election-day performance was *better than his polling on average in every one of the 10,000 simulations*. This is yet more evidence that the Wilder effect, strong in the early 1990s, is strong no longer.¹⁶ In light of the paper’s findings on female candidates, the appropriate conclusion is not that Senator Clinton suffered from a Whitman effect, but that turnout, survey sampling, or campaign effects explain the gap we observe.

4 Conclusion

The Wilder effect occupies an unusual position in our thinking about American elections, as it is often invoked (e.g. Elder, 2007; Lanning, 2005) but rarely scrutinized. By analyzing Senate and Gubernatorial elections between 1989 and 2006, this paper has provided the first large-sample test of the Wilder effect. In the early 1990s, there was a pronounced gap between polling and performance for black candidates of about 2.3 percentage points. But in the mid-1990s, that upward bias in telephone surveys disappeared. At a time when scholars are increasingly

¹⁵Florida and Michigan were not sites of active campaigning, and Obama did not appear on the ballot in the latter, so both are excluded.

¹⁶In assessing this finding, one should also see Greenwald (2008), which provides evidence of heterogeneity across states.

concerned about the validity of phone surveys, these results provide some reassurance. We have also seen that the polling-performance gap is closely related to a candidate’s level of pre-election support, meaning that we should not naively attribute the entire observed gap in a given election to racial bias. Douglas Wilder, David Dinkins, and Tom Bradley were all front-runners, and so all could have expected a small decline in their election day performance. Even female front-runners should expect declines into election day, although they are not subject to any Wilder-style bias.

Our ability to isolate the mechanisms underpinning these results is limited by the aggregated nature of the data. Future work on the micro-level mechanisms of the Wilder effect might compare outcomes to exit polls, since there are fewer sources of bias when measuring reported vote choice just moments after the vote itself. Still, as Traugott and Price (1991) note, the exit polls available for the elections discussed here have been re-weighted to match the actual election outcomes, making it impossible to use publicly available exit polls to identify biases. Another important extension is to test whether Latino and Asian American candidates face a polling-performance gap, or whether they ever did. From what we have seen here, the Wilder effect is a particular product of egalitarian norms, anti-black biases, and a specific period in political history, but testing additional groups will lend added certainty to those conclusions.

From the available data, we cannot say definitively why the Wilder effect for black candidates disappeared. It is possible that shifts in black leadership styles played a role, although the candidates most closely associated with the Wilder effect—including Wilder, Bradley, and Dinkins—all de-emphasized race and focused on mobilizing white support (Hajnal, 2007). Still, the observed patterns do make some explanations more likely than others. The Wilder effect was present in the early 1990s, in the north as well as the south, and in campaigns that were heavily racialized as well as those that were not.¹⁷ It disappeared swiftly at about the time that welfare reform silenced one critical, racialized issue. And even when particular campaigns invoked racialized appeals after that, the effect did not reappear. This lends credence to the

¹⁷Harvey Gantt’s 1990 run for Senate against Jesse Helms, in which he faced an ad about racial quotas taking whites’ jobs, is a clear case of the former, while Alan Keyes’ 1992 run for Senate is an example of the latter.

claim that the salience of racialized issues nationally facilitates Wilder effects locally. It also opens up the possibility that the Wilder effect could return if race again infuses prominent topics in national politics.

References

- Achen, Christopher. 2005. "Let's Put Garbage-Can Regressions and Probits Where They Belong." *Conflict Management and Peace Research* 22:327–339.
- Alexander, Deborah and Kristi Andersen. 1993. "Gender as a Factor in the Attribution of Leadership Traits." *Political Research Quarterly* 46(3):527.
- Aquilino, William S. 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment." *Public Opinion Quarterly* 58(2):210–240.
- Baumgartner, Frank R., Bryan D. Jones, John Wilkerson and E. Scott Adler. 2006. "The Policy Agendas Project." Available online at www.policyagendas.org.
- Beck, Nathaniel, Simon Jackman and Howard Rosenthal. 2006. "Presidential Approval: The Case of George W. Bush." Paper presented at the annual meeting of the Society for Political Methodology, Davis, California.
- Berinsky, Adam J. 1999. "The Two Faces of Public Opinion." *American Journal of Political Science* 43(4):1209–30.
- Berinsky, Adam J. 2004. "Can We Talk? Self-Presentation and the Survey Response." *Political Psychology* 25(4):643–659.
- Bishop, George F. and Bonnie S. Fisher. 1991. "Secret Ballots and Self-Reports in an Exit Poll Experiment." *Public Opinion Quarterly* 59(4):568–588.
- Citrin, Jack, Donald P. Green and David O. Sears. 1990. "White Reactions to Black Candidates: When Does Race Matter?" *Public Opinion Quarterly* 54:74–96.
- Davis, C.E. 1976. "The Effect of Regression to the Mean in Epidemiologic and Clinical Studies." *American Journal of Epidemiology* 104(5):493–498.
- Elder, Janet. 2007. "Will There Be an Obama Effect?" *New York Times* (16 May).
- Erikson, Robert S. and Christopher Wlezien. 2007. "Are Political Markets Really Superior to Polls as Election Predictors?" Draft.
- Finkel, Steven E., Thomas M. Guterbock and Marian J. Borg. 1991. "Race-of-Interviewer Effects in a Pre-Election Poll." *Public Opinion Quarterly* 55(3):313–330.
- Fowler, Floyd J., Anthony M. Roman and Zhu Xiao Di. 1998. "Mode Effects in a Survey of Medicare Prostate Surgery Patients." *Public Opinion Quarterly* 62(1):29–46.
- Frey, David L. and Samuel L. Gaertner. 1986. "Helping and the avoidance of inappropriate interracial behavior: a strategy that perpetuates a nonprejudiced self-image." *Journal of Personality and Social Psychology* 50(6):1083–1090.
- Gilens, Martin. 1999. *Why Americans Hate Welfare: Race, Media, and the Politics of Anti-poverty Policy*. Chicago, IL: University of Chicago Press.
- Greenwald, Anthony G. 2008. "Tracking the Race Factor." Pew Research Center, available online at: <http://pewresearch.org>.
- Hajnal, Zoltan L. 2007. *Changing White Attitudes toward Black Political Leadership*. New York: Cambridge University Press.

- Highton, Benjamin. 2004. "White Voters and African American Candidates for Congress." *Political Behavior* 26(1):1–25.
- Huber, Gregory A. and John S. Lapinski. 2006. "The Race Card Revisited: Assessing Racial Priming in Policy Contests." *American Journal of Political Science* 50(2):421–440.
- Huddy, Leonie and Nayda Terkildsen. 1993. "The Consequences of Gender Stereotypes for Women Candidates at Different Levels and Types of Office." *Political Research Quarterly* 46(3):503.
- Hurwitz, Jon and Mark Peffley. 1997. "Public Perceptions of Race and Crime: The Role of Racial Stereotypes." *American Journal of Political Science* 41(2):375–401.
- Jackson, Herb. 1993. "Whitman Rips Pollsters Who Failed to Predict Her Victory." *Associated Press* (November 3).
- Jamieson, Kathleen Hall. 1995. *Beyond the Double Bind: Women and Leadership*. New York, NY: Oxford University Press.
- Jones, Bryan D. and Frank R. Baumgartner. 2005. *The Politics of Attention: How Government Prioritizes Problems*. Chicago, IL: University of Chicago Press.
- Kahn, Kim F. 1994. "Does gender make a difference? An experimental examination of sex stereotypes and press patterns in statewide campaigns." *American Journal of Political Science* 38(1):162–195.
- Kahn, Kim F. 1996. *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. New York: Columbia University Press.
- Keeter, Scott and Nilanthi Samaranyake. 2007. "Can You Trust What Polls Say about Obama's Electoral Prospects?" Pew Research Center Report, available online at: <http://pewresearch.org>.
- Kinder, Donald R. and David O. Sears. 1981. "Prejudice and Politics: Symbolic Racism Versus Racial Threats to the Good Life." *Journal of Personality and Social Psychology* 40:414–31.
- King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.
- Koch, Jeffrey M. 2002. "Gender Stereotypes and Citizens' Impressions of House Candidates' Ideological Orientation." *American Journal of Political Science* 46:453–462.
- Krysan, Maria. 1998. "Privacy and the Expression of White Racial Attitudes: A Comparison across Three Contexts." *Public Opinion Quarterly* 62(4):506–544.
- Lanning, Kevin. 2005. "The Social Psychology of the 2004 U.S. Presidential Election." *Analyses of Social Issues and Public Policy* 5(1):145–152.
- McDonald, Michael P. and Samuel L. Popkin. 2002. "The Myth of the Vanishing Voter." *American Political Science Review* 94(4):963–974.
- Melton, R.H. and Richard Morin. 1989. "Wilder Taking Command in Va. Race, Polls Show." *The Washington Post* (October 29).
- Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, NJ: Princeton University Press.
- Reeves, Keith. 1997. *Voting Hopes or Fears? White Voters, Black Candidates, and Racial Politics in America*. New York, NY: Oxford University Press.
- Rice, John A. 1995. *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press.
- Rosenwasser, Shirley, Robyn Rogers, Sheila Fling, Kayla Silvers-Pickens and John Butemeyer.

1987. "Attitudes toward women and men in politics: Perceived male and female candidate competencies and participant personality characteristics." *Political Psychology* 8(2):191–200.
- Sanbonmatsu, Kira. 2002. "Gender Stereotypes and Vote Choice." *American Journal of Political Science* 46(1):20–34.
- Sapiro, Virginia. 1981. "If US Senator Baker were a woman: An experimental study of candidate images." *Political Psychology* 2:61–83.
- Schafer, Joseph L. 1997. *Analysis of incomplete multivariate data*. London: Chapman & Hall.
- Schuman, Howard, Charlotte Steeh, Lawrence Bobo and Maria Krysan. 1997. *Racial Attitudes in America: Trends and Interpretations*. Cambridge, MA: Harvard University Press.
- Sears, David O., John J. Hetts, Jim Sidanius and Lawrence Bobo. 2000. Race in American Politics. In *Racialized Politics: The Debate about Racism in America*, ed. David O. Sears, Jim Sidanius and Lawrence Bobo. University of Chicago Press.
- Sears, David O., P.J. Henry and Rick Kosterman. 2000. Egalitarian Values and Contemporary Racial Politics. In *Racialized Politics: The Debate about Racism in America*, ed. David O. Sears, Jim Sidanius and Lawrence Bobo. University of Chicago Press.
- Sniderman, Paul and Edward Carmines. 1997. *Reaching Beyond Race*. Cambridge, MA: Harvard University Press.
- Sniderman, Paul M., Gretchen C. Crosby and William G. Howell. 2000. The Politics of Race. In *Racialized Politics: The Debate about Racism in America*, ed. David O. Sears, Jim Sidanius and Lawrence Bobo. University of Chicago Press.
- Streb, Matthew J., Barbara Burrell, Brian Frederick and Michael A. Genovese. 2008. "Social Desirability Effects and Support for a Female American President." *Public Opinion Quarterly* 72(1):76–89.
- Traugott, Michael W. and Vincent Price. 1991. "A Review: Exit Polls in the 1989 Virginia Gubernatorial Election." *Public Opinion Quarterly* 56(2):245–253.
- Valentino, Nicholas. 1999. "Crime and the Priming of Racial Attitudes During Evaluations of the President." *Public Opinion Quarterly* 63:293–320.
- White, Ismail K. 2007. "When Race Matters and When It Doesn't: Racial Group Differences in Response to Racial Cues." *American Political Science Review* 101(02):339–354.
- Wlezien, Christopher and Robert S. Erikson. 2002. "The Timeline of Presidential Election Campaigns." *Journal of Politics* 64:969–993.
- Wooldridge, Jeffrey M. 2003. "Cluster-Sample Methods in Applied Econometrics." *American Economic Review* 93(2):133–138.

	Min.	Max.	Mean	SD
Poll	0.29	0.66	0.45	0.10
Outcome	0.29	0.61	0.44	0.08
Gap	-0.06	0.09	0.01	0.04
Female	0.00	1.00	0.13	0.34
Year	1989	2006	1998	6
Democrat	0.00	1.00	0.72	0.45
Governor	0.00	1.00	0.40	0.50
Is Incumbent	0.00	1.00	0.06	0.25
Against Incumbent	0.00	1.00	0.43	0.50
In South	0.00	1.00	0.36	0.49
Pct. Black in State	0.04	0.33	0.17	0.08
State Share Dem	-0.11	0.14	0.02	0.07
Turnout	0.34	0.62	0.45	0.07
Δ Turnout	-0.06	0.10	0.03	0.05
Days Until Election	1.00	25.00	10.47	6.95
Poll's Sample Size	400	1150	739	203

Table 1: Key variables for the 47 pre-election polls observed in 18 races for Senate or Governor involving African Americans

	Min.	Max.	Mean	SD
Poll	0.25	0.74	0.49	0.11
Outcome	0.28	0.71	0.49	0.08
Gap	-0.19	0.10	-0.00	0.05
Female	1.00	1.00	1.00	0.00
Year	1990	2006	1998	5
Democrat	0.00	1.00	0.72	0.45
Governor	0.00	1.00	0.46	0.50
Is Incumbent	0.00	1.00	0.22	0.42
Against Incumbent	0.00	1.00	0.33	0.47
In South	0.00	1.00	0.13	0.34
Pct. Black in State	0.00	0.33	0.10	0.08
State Share Dem	-0.15	0.15	0.01	0.06
Turnout	0.32	0.70	0.51	0.10
Δ Turnout	-0.09	0.12	0.02	0.05
Days Until Election	0.00	33.00	11.53	7.76
Poll's Sample Size	118	2326	715	325

Table 2: Key variables for the 208 pre-election polls observed in 119 races for Senate or Governor involving females

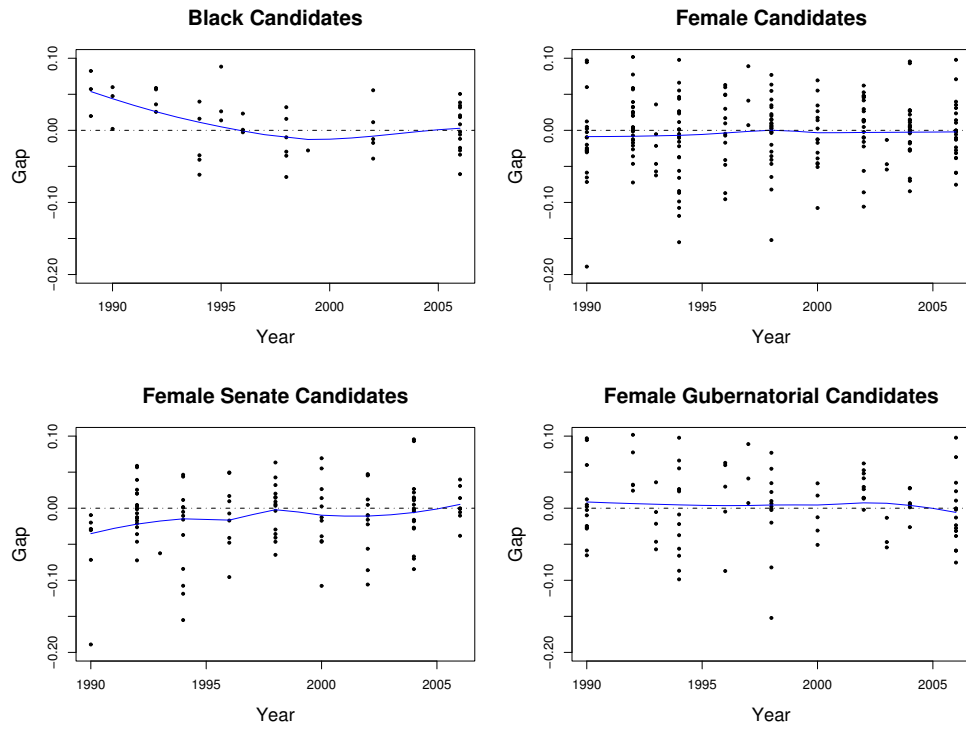


Figure 1: These figures depict the distribution of the polling-performance gap by year, and also providing loess smoothing lines.

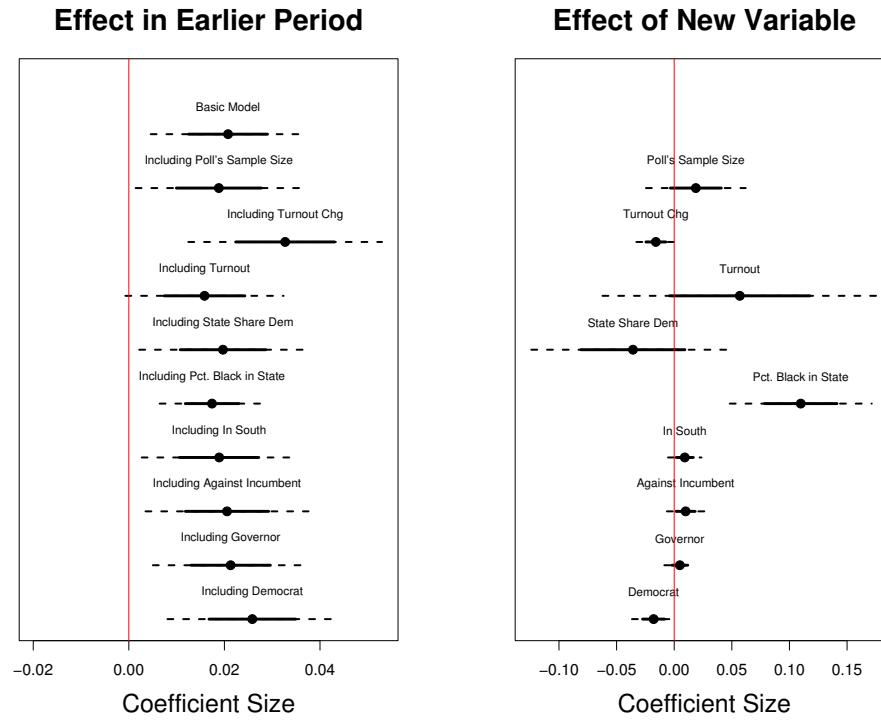


Figure 2: The figure at left shows the coefficient for elections occurring before 1996. It demonstrates the robustness of the finding that elections before 1996 had a larger Wilder effect. At right, we see the coefficient and uncertainty for the newly included variable in each model.

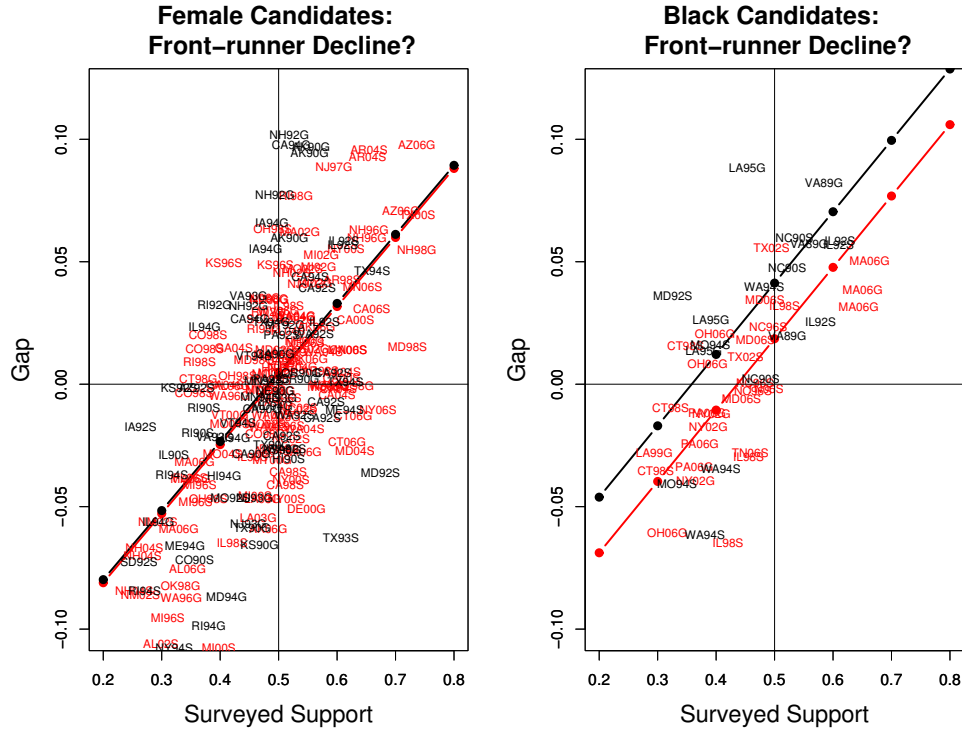


Figure 3: These figures show that for both female (at left) and black candidates, the polling-performance gap is strongly related to the candidate's initial level of support. Statistically, the slopes of these lines are indistinguishable.

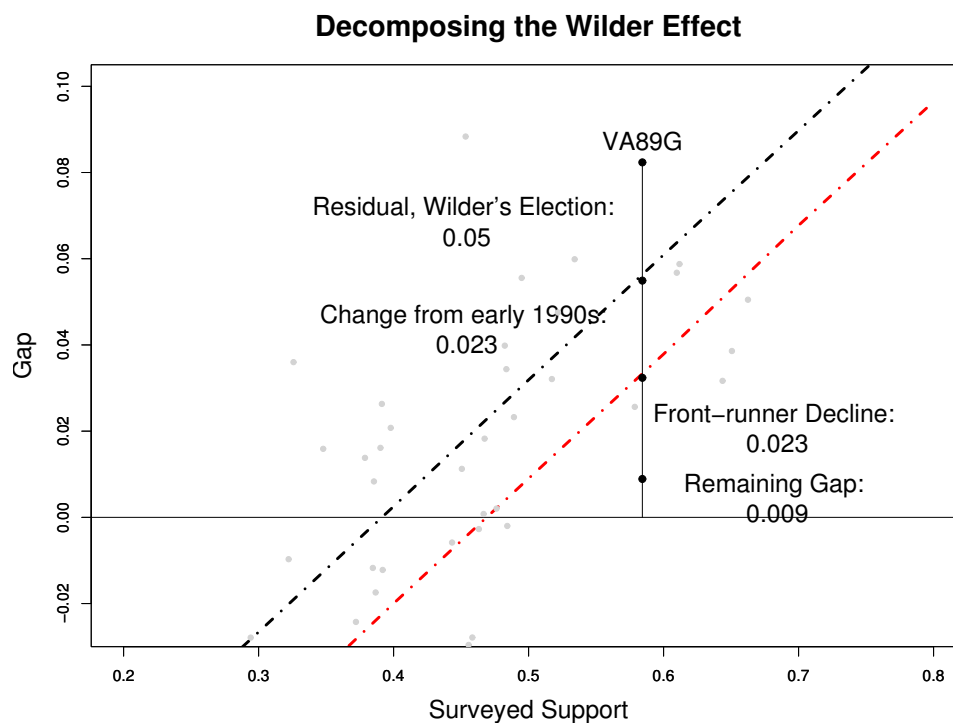


Figure 4: Here, we decompose the largest observed polling-performance gap for Douglas Wilder's 1989 Gubernatorial bid using predictions from a linear model. 2.3 percentage points of the 8.2 percentage point gap are attributable to the expected decline for a front-runner polling at 58%.