

NBER WORKING PAPER SERIES

IDENTIFYING CONFIRMATORY BIAS IN THE FIELD:  
EVIDENCE FROM A POLL OF EXPERTS

Rodney J. Andrews  
Trevon D. Logan  
Michael J. Sinkey

Working Paper 18064  
<http://www.nber.org/papers/w18064>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2012

We thank Dann Arce, Katie Baldiga, Chetan Dave, Seth Freedman, Nathan Fritz-Joseph, P.J. Healy, Pok-sang Lam, Brendan Nyhan, Dan Stone, seminar audiences at Mississippi State, Ohio State, and participants at the SEA meetings for helpful comments. Kyle Kain, Michael Kovach and Chris Zukauckas provided excellent research assistance. The usual disclaimer applies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Rodney J. Andrews, Trevon D. Logan, and Michael J. Sinkey. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Identifying Confirmatory Bias in the Field: Evidence from a Poll of Experts  
Rodney J. Andrews, Trevon D. Logan, and Michael J. Sinkey  
NBER Working Paper No. 18064  
May 2012  
JEL No. D01,D03,N32

### **ABSTRACT**

Laboratory experiments have established the existence of cognitive biases, but their explanatory power in real-world economic settings has been difficult to measure. We estimate the extent of a cognitive bias, confirmatory bias, among experts in a real-world environment. In the Associated Press Top 25 College Football Poll expert pollsters are tasked with assessing team quality, and their beliefs are treated week-to-week with game results that serve as signals about an individual team's quality. We exploit the variation provided by actual game results relative to market expectations to develop a novel regression-discontinuity approach to identify confirmatory bias in this real-world setting. We construct a unique personally-assembled dataset that matches more than twenty years of individual game characteristics to poll results and betting market information, and show that teams that slightly exceed and barely miss market expectations are exchangeable. The likelihood of winning the game, the average number of points scored by teams and their opponents, and even the average week of the season are no different between teams that slightly exceed and barely miss market expectations. Pollsters, however, significantly upgrade their beliefs about a team's quality when a team slightly exceeds market expectations. The effects are sizeable-- nearly half of the voters in the poll rank a team one slot higher when they slightly exceed market expectations; one-fifth of the standard deviation in poll points in a given week can be attributed to confirmatory bias. This type of updating suggests that even when informed agents make repeated decisions they may act in a manner which is consistent with confirmatory bias.

Rodney J. Andrews  
The University of Texas at Dallas  
800 West Campbell Road  
MS WT21  
Richardson, TX 75080  
and NBER  
rodney.j.andrews@utdallas.edu

Michael J. Sinkey  
University of West Georgia  
Richards College of Business  
Carrollton, GA 30118  
msinkey@westga.edu

Trevon D. Logan  
The Ohio State University  
410 Arps Hall  
1945 N. High Street  
Columbus, OH 43210  
and NBER  
logan.155@osu.edu

# 1 Introduction

Despite the increasing evidence that there are behavioral and cognitive biases in decision making, the general relevance of such effects to real-world economic phenomena has been questioned. Decision making in the market is inherently different from the controlled environment of the laboratory. Many biases found in laboratory settings may be attributable to relatively uninformed individuals making unfamiliar or irregular decisions. While individuals may have idiosyncratic biases, market aggregation and experience may limit the scope of such effects. In general, it has been difficult to establish that experienced decision-makers consistently exhibit cognitive or behavioral biases, or that the market effects are large (Levitt and List, 2008). As such, laboratory estimates of bias may have little explanatory power in real-world settings (Al-Ubaydli and List, 2012).

In this paper we estimate the effects of a cognitive bias among experts in a real-world setting. Specifically, we examine the extent to which confirmatory bias affects announced expert opinions. Confirmatory bias is defined as “the use of weak evidence to bolster an existing hypothesis” (Rabin and Schrag, 1999). While examples abound of instances where individuals over-infer from weak evidence when it is consistent with their existing hypotheses, it could be the case that confirmatory bias is cancelled out by other biases so that the net impact on belief updating is small. Although confirmatory bias has been documented in laboratory settings, there is little field evidence that connects this cognitive bias to economic outcomes or measures the degree to which it influences outcomes.

We exploit a novel real-world setting to identify the magnitude of confirmatory bias in announced expert opinions: the Associated Press Top 25 College Football Poll, a weekly poll that elicits expert opinions about the quality of college football teams. The Associated Press College Football Poll presents an ideal environment to identify confirmatory bias outside of the laboratory. The poll consists of a panel of experts who have incentives to rank teams truthfully according to their preferences. Voters are neither compensated nor rewarded for favoring certain teams over others. The AP poll differs substantially from the Coaches’ Poll, where there may be explicit conflicts of interest in voting (Kotchen and Potoski, 2011). Because of this, we have few of the concerns that frequently plague empirical studies of behavioral biases, such as the potential of

subjects misunderstanding the task, individuals seeking to pursue other tasks rather than those designated, or individuals responding to concerns over their reputation in a manner which would elicit less-than-truthful rankings (Sinkey, 2011). In addition, the stakes in the poll are reasonably high. NCAA football plays a large role in the national landscape. Since more than a quarter of the U.S. population closely follows college football (Kotchen and Potoski, 2011), the opinions of these experts are watched very closely. An additional strength is that the poll allows us to focus on beliefs and belief formation outside of the laboratory, helping us build stronger links between the laboratory and the field (Al-Ubaydli and List, 2012).

As there are distinct advantages to explicitly modeling the setting in field experiments (Card et al., 2011), we employ a theoretically-motivated estimation strategy to provide the first field evidence for confirmatory bias. We model confirmatory bias as a previous hypothesis-driven response to noisy information. Individuals construct beliefs about the value of a parameter—in this case, the quality of a collegiate football team—from information in their environment using Bayes’ rule and then announce those beliefs. In our model, individuals explicitly replace ambiguous information with confirmatory information. For example, when they believe that the value of the parameter is positive they replace weak evidence for a positive value with information that is stronger evidence for a positive value. We model the magnitude of this replacement as confirmatory bias.

Our identification strategy builds on the central insight from models of confirmatory bias: since individuals with confirmatory bias over-infer from weak signals, the point at which information changes from being weakly negative to weakly positive serves as the trigger point for confirmatory bias. Standard models of Bayesian updating predict that individuals would upgrade or downgrade their beliefs with tempered and smooth updating when confronted with small differences from expectations. In contrast, models of confirmatory bias predict that biased individuals would use these same small differences to markedly and discontinuously change their beliefs and thus confirm their prior hypotheses. By making use of quasi-experimental techniques, we are able to isolate the effect of confirmatory bias from the effects of other cognitive biases, such as the primacy effect.

When confirmatory bias is present the replacement of information in either direction changes discontinuously at the point where weakly negative information becomes weakly positive information.

While agents can correctly distinguish between positive and negative signals, they over-infer from weakly positive or weakly negative signals that conform to their prior beliefs. Put another way, individuals can tell good from bad, but those with confirmatory bias replace “weak” good with “strong” good. This intuitive theoretical insight motivates our use of a regression-discontinuity (RD) approach to estimating confirmatory bias.

We use betting lines as a proxy for market expectations, which are particularly attractive because they serve as summaries of the information available prior to games.<sup>1</sup> The poll ranks the teams thought to be the 25 best teams in the country during a given week. Beliefs of poll voters are treated weekly with signals about the quality of college football teams, and we match more than twenty years of poll data to market expectations of game results, betting lines, and a rich set of game data which serve as public signals available to voters.<sup>2</sup>

We use the difference between expected and actual margin of victory as the information that discontinuously changes from being weakly negative to weakly positive. We hypothesize that small differences between the expected and realized margin of victory are ambiguous signals about a team’s quality. When a given team performs very close to market expectations, very little new information about the team’s quality is revealed. Indeed, by definition, the team has done about as well as expected to when the predicted and actual margin of victory are close. Furthermore, the scoring in football makes it difficult for teams to manipulate margins of victory by very small amounts. Most important for our research design, poll voters are unable to manipulate the betting line, final score or margin of victory in a given contest.

Our unique data allow us to demonstrate that a number of important characteristics—whether or not the team won the game, the number of points that a team scores, the number of points the team’s opponent scores, the quality of the opponent, the location of the game, and even the week of the season—do not change discontinuously when the difference between the actual margin of victory and the final betting line is small. These determinants are neither substantively nor statistically

---

<sup>1</sup>Following Card and Dahl (2011) we assume that betting lines are sufficient to capture the expected outcome of the game. Logan and Sinkey (2011) and Sinkey (2011) show that betting lines are extremely accurate predictions of actual margins of victory in college football.

<sup>2</sup>There is ample qualitative evidence to support the claim that sportswriters look at betting lines as salient sources of information about team qualities. For example, sportscasters who are voters in the AP poll frequently mention the betting lines during television broadcasts.

different between teams that either just exceed or just miss exceeding the expected margin of victory. This finding corroborates the assumption that little, if any, substantive information is revealed when the difference between the actual margin of victory and the final betting line is small, allowing us to treat teams who barely exceed or barely miss exceeding the expected margin of victory as exchangeable.

We concentrate on a small window where our signal of interest is changing but where other signals do not which permits us to isolating the response to this signal. Since the other observable characteristics are indistinguishable on either side of the discontinuity, we avoid the problem of having to specify the full updating function to produce the counterfactual. One empirical hurdle in the literature is that agents may respond with different types of bias to different signals, making it difficult to draw inference about the magnitude of any particular bias. Because all the public information that experts use is transparent or known and all other observable signals are the same on either side of the discontinuity, there is no other plausible updating in beliefs in response to a signal other than the one we seek to identify.

We use the point where the betting line is exactly equal to the actual margin of victory as the discontinuity and focus on a very small window around that point. We interpret the presence of sharp discontinuities in belief updating as evidence of confirmatory bias. We find that these expert pollsters react significantly to weakly positive signals even though there are no substantive differences on either side of the discontinuity. In our main specification, we find that roughly half of the poll voters increase a team's ranking by one slot due to confirmatory bias. Our estimates suggest that around one-fifth of the overall standard deviation in poll points can be attributed to this effect. These results are in sharp contrast to standard models of Bayesian updating, which would have both smooth and symmetric updating through the discontinuity. Furthermore, we show that the amount of confirmatory bias we estimate does not dissipate over time. Experts react similarly to the weak signals no matter when they are received. Experience *within* a season does not alter the bias, as individuals do not "learn away" the bias by receiving more information about a team's quality. Our results imply that confirmatory bias is prevalent in the Associated Press Top 25 College Football Poll, and the magnitude is economically significant.

## 2 Identifying Confirmatory Bias

### 2.1 Confirmatory Bias and Bounded Rationality

Confirmatory bias has been well-documented in the psychology literature—examples include Mahoney (1977); Lord, Ross, and Lepper (1979); Darley and Gross (1983); Nickerson (1998). It is now well-known that individuals have tendencies to seek out information that confirms their prior hypotheses (Jones and Sudgen (2001); Meissner and Kassin (2002)). Theoretically, modeling confirmatory bias requires a dynamic environment, since individuals first form hypotheses and then use their hypotheses as guides for interpreting new information. Confirmatory bias also requires a degree of signal variation. Agents must respond differently to different sorts of signals, such as over-infering from weak signals.

The most well-known model of confirmatory bias in economics comes from Rabin and Schrag (1999), who model confirmatory bias in a simplified environment with two signals that correspond to two possible states of the world. In their model, individuals are prone to stark misinterpretations of signals as they may incorrectly interpret binary signals as being consistent with previous hypotheses of the outcomes. These misinterpretations then strengthen the hypothesis, generating persistence of the bias over time. Wilson (2003) frames confirmatory bias as an optimal decision under bounded memory, where decision makers have a limited amount of ‘states’ that they can remember and optimally choose to remember states in a way that mirrors confirmatory bias. Yariv (2005) models confirmatory bias as an extension of belief utility, where individuals receive utility out of directional confidence, or having beliefs that support similar actions as previous beliefs. Gottlieb (2010) frames confirmatory bias as a result of selective awareness: once an individual becomes convinced of a certain state of the world it becomes prohibitively costly to consider contrary evidence. Other theoretical models concentrate on establishing how confirmatory bias and other biases could arise from bounded rationality (Gennaioli and Shleifer (2010), among others).<sup>3</sup>

While these models provide motivation for why rational individuals might conduct biased inference, they are silent on whether these biases are sizeable or how pervasive they are. We are

---

<sup>3</sup>Conditional on the existence of confirmatory bias, the models say little about the magnitude of confirmation bias.

agnostic about the specific source of confirmatory bias. Instead, we concentrate on the empirical predictions shared by these models. We take from these models the central insight that confirmatory bias affects the ways that new information is interpreted, and that confirmatory bias can be defined as an over-reaction to weak evidence which corroborates prior hypotheses. We use this as our definition of confirmatory bias in this paper, but note that these models classify other types of inference and signal response as confirmatory bias which we do not attempt to estimate.<sup>4</sup>

## 2.2 Distinguishing Confirmatory Bias From Other Cognitive Biases

Confirmatory bias and other psychological or cognitive biases may be observationally equivalent. If the change in beliefs that we find is due to either another cognitive bias or some mixture of another bias and confirmatory bias, then our estimates of confirmatory bias are biased. Our approach allows us to distinguish confirmatory bias from other observationally equivalent cognitive biases: availability heuristics, primacy effects, and Bayesian over/underreaction. Below, we describe how our approach identifies confirmatory bias as opposed to these other possibilities.

Biases based on the availability heuristic occur when individuals conduct inference based on whatever information can be brought to mind, which may be correlated with confirmatory bias if the only information a person can remember is their previous hypothesis. Empirically, our data and estimation strategy allow us to distinguish confirmatory bias from the availability heuristic. In each week of the poll, only one game for each team is available for voters to consider, and thus, we are able to reduce and accurately measure the new information that voters have to consider when updating their beliefs. In our environment, the amount of information available and the amount of information that voters consider are proper subsets of each other. As such, our results are not driven by individuals discarding information.<sup>5</sup>

Similarly, the over-weighing of information in confirmatory bias may also be mechanically similar to the primacy effect, where individuals only remember initial information. Confirmatory bias differs from primacy effects in that primacy supposes that individuals are selective about which

---

<sup>4</sup>For example, confirmatory bias could arise when agents over-infer from ambiguous information, which would be positive or negative.

<sup>5</sup>We do not assume that individuals weigh all information equally.



information is considered based on when it arrived, but unselective in how information is interpreted, while confirmatory bias is unselective about which information is considered based on when it arrived, but selective in how it is interpreted. We would not expect individuals to react to new information at all with the primacy effect, and our estimation strategy explicitly estimates responses to new information. By design, we limit the possibility that the primacy effect is driving our results.

Bayesian underreaction occurs when individuals do not sufficiently revise their priors based on new information. Our approach identifies belief updating that would not be driven by Bayesian underreaction. Underreaction to all information is inherently different from overreactions to a particular type of information. If underreaction was present, then it would still be the case that agents would react similarly to all types of signals. The agents' responses would be dampened by Bayesian underreaction in all cases. Since we estimate different reactions to different types of information, our results would not be due to underreaction. The same applies to Bayesian overreaction.

Although not a cognitive bias, herd behavior is a potential explanation for our results. Since we use poll evidence, it could be the case that voter responses are due to herd behavior among poll voters. Our strategy also allows us to distinguish confirmatory bias from herd behavior. In models of herd behavior, agents look at the previous decisions of peers when forming their own opinions as in Banerjee (1992). If the previous decisions of other voters in the poll align with a given voter's hypothesis, then herd behavior would be observationally equivalent to confirmatory bias. We are able to distinguish confirmatory bias from herd behavior by looking at changes in the poll in response to a new, ambiguous signal. Herding requires a voter to know how voters respond because they believe that the opinions of others have information. We look at voting in response to a specific signal, and voters vote simultaneously without knowledge of how others have responded to the same signal. Any herding observed would have to be a function of subsequent voting, not the contemporaneous voting we exploit. Put another way, herding requires that voters know how others have voted, and since the votes of others are not known at the time of voting herd behavior cannot explain the changes in beliefs we observe.

### 2.3 Advantages of the Regression Discontinuity Approach

We use a quasi-experimental technique to identify a cognitive bias in the field. Our novel approach differs from the ways that others have attempted to identify cognitive biases in observational data. For example, other studies have attempted to identify overall updating in the AP poll (Stone (2011), Stone and Zafar (2010)). This strategy requires researchers to simulate how voters would respond to all potential signals. Those simulations are then used as counterfactuals to estimate updating. However, beliefs cannot be perfectly simulated because the information that individuals use to form beliefs can never be perfectly simulated. Therefore, belief simulation offers limited insight toward identifying cognitive biases. The inability to simulate beliefs with a high degree of certainty poses a problem for identification of biases since one cannot distinguish cognitive biases in updating from weaknesses of the simulation as the simulations impose strong assumptions about belief formation that are difficult to corroborate.

Our approach circumvents this problem by first modeling how agents would respond to information if they had confirmatory bias. In our model, confirmatory bias—and not other cognitive biases—is triggered discontinuously by the receipt of different weak signals. We are only required to plausibly model the structure for the responses of beliefs to this specific information. Unlike other studies, we do not attempt to simulate beliefs and use those simulations as our counterfactuals for rational updating. We can explicitly measure both signals and responses in our quasi-experimental setting. As such, we isolate a local response to a specific signal and do not make claims to identify overall updating given the complications involved in deriving a justifiable counterfactual.

We exploit expectations about game results by using the difference between the actual margin of victory and the final betting lines to summarize the information about the quality of a given team that can be gleaned from a given contest. This allows us to construct both weak and strong signals of information, key for identifying confirmatory bias. Using the difference between the predicted and actual margin of victory allows us to define outcomes that would and would *not* provide plausibly new information about quality.

We concentrate on the set of contests where outcomes closely match expectations, a subset of contests that provides weak signals of quality in both directions but where (we show later) other

substantive measures of quality do not differ. As these other substantive indicators of quality do not differ in this window, we avoid the problems of having to specify the full updating function to provide a counterfactual. Since the only variation is with respect to the signal provided by outcomes relative to expectations, our counterfactual is the intuitive notion that Bayesian updating would respond smoothly and symmetrically to this signal, while confirmatory bias would change discontinuously at the point where the signal changes from being weakly negative to being weakly positive. We believe that this approach is a more defensible empirical strategy to identify cognitive biases in observational data. Furthermore, using the RD approach focuses our attention on the set of contests where confirmatory bias would be present but where other biases would not produce similar irregularities.

### 3 Modeling Confirmatory Bias

Using previous models of confirmatory bias as a guide, we model confirmatory bias as a dynamic response to signals in the context of a signal extraction problem. Like Rabin and Schrag (1999), our model is a reduced-form model of confirmatory bias. Our model highlights the key empirical prediction of confirmatory bias we test and makes clear why the regression discontinuity design is an appropriate test of the existence and magnitude of confirmatory bias. We model individuals as decision makers who update their beliefs about the distribution of a parameter that evolves over time. Individuals form hypotheses before receiving information and interpret the information with respect to the value of the parameter.

Specifically, there are  $n$  parameters of interest which correspond to quality assessments for each team  $i$ ,  $q_{i,t}, i \in \{1, 2, \dots, n\}$  which we assume evolve according to random walks with unknown drifts, i.e.,  $q_{it} = q_{i,t-1} + \mu_{i,t}$ . In every period  $t$ , individuals receive a noisy signal about the value of the drift term for each parameter,  $\theta_{i,t}$ , such that  $\theta_{i,t} = \mu_{i,t} + \epsilon_{i,t}$ , where  $\epsilon_{i,t} \sim N(0, \sigma_\epsilon^2)$ , where the shocks are both i.i.d. and uncorrelated both across time and between parameters, i.e.,  $\epsilon_{i,t} \perp \epsilon_{-i,t}$  and  $\epsilon_{i,t} \perp \epsilon_{i,t-i}$ . We make the assumption that individuals have a common prior about the distribution for  $\mu_{i,0}$ :  $\mu_{i,0} \sim N(0, \sigma_{\mu_{i,0}}^2)$  and that agents have a common prior about the initial level of quality,

$q_{i,0}$ .<sup>6</sup>

Observers use Bayes' rule and engage in Bayesian updating about both the value of the quality parameter for each team and the moments of the distribution for the drift term. In our setting, the drift term represents how much a team has either improved or regressed in each week, while the distribution for the drift term represents the potential capacity to improve or decline. This distribution for the drift term is fixed at the start of time; however, every period individuals view a signal related to the week-specific draw from the distribution (which dictates how much a team has improved or worsened in quality in that given week), and they use those signals to pin down the parameters of the probability distribution.

We concentrate on one noisy signal  $\theta_{i,t}$ . While quality is a function of several informative signals, we model the responses of individuals to  $\theta_{i,t}$ , as it is the only source of variation in our empirical model. We assume that contests that are fairly close to the threshold are exchangeable; the RD design allows us to present evidence that corroborates this assumption. Exchangeability implies that the other signals do not vary meaningfully, and therefore, responses to the other signals are unrelated to the response to the signal of interest.

As we exploit the exchangeability assumption of the RD design, the other signals do not vary and therefore responses to them are unrelated to responses to the signal of interest.

For simplicity, we model agents as minimizing their mean-squared error using an additively separable quadratic loss function for agent utility based on the veracity of predictions  $\hat{q}_{i,t}$  compared to truthful predictions:<sup>7</sup>

$$U^E = \sum_{i=1}^n -(\hat{q}_{i,t} - q_{i,t})^2. \quad (1)$$

Modeling agents in this fashion rules out an interpretation of confirmatory bias as selective use of information, since selecting information to match a hypothesis would not be utility-maximizing. After the first period, the belief about the mean of the distribution of  $\mu$  given the realization of  $\theta$

---

<sup>6</sup>Practically speaking, the initial prior belief about team quality corresponds to the preseason AP poll.

<sup>7</sup>Other utility functions, such as least absolute deviation, would also induce truthful reporting.

is utility-maximizing when:

$$\mu_{i,1}^{\hat{}} = \mathbb{E}(\mu_{i,1}|\theta_{i,1}) = \frac{f(\mu)f(\theta|\mu)}{\int_{-\infty}^{\infty} f(\mu)f(\theta|\mu)\partial\mu} = \frac{\sigma_{\epsilon}^2}{\sigma_{\mu,0}^2 + \sigma_{\epsilon}^2}\theta_{i,1} \quad (2)$$

with variance:

$$\widehat{\sigma_{\mu,1}^2} = Var(\mu|\mu_{i,1}^{\hat{}}) = \frac{\sigma_{\epsilon}^2\sigma_{\mu,0}^2}{\sigma_{\mu,0}^2 + \sigma_{\epsilon}^2}. \quad (3)$$

Observers make announcements about the quality parameter  $q_{i,t}$  every period after they update their beliefs about the drift of  $q$ . In general, without the presence of confirmatory bias, the estimate for  $\mu_i$  in any period  $t$  would be:

$$\mu_{i,t}^{\hat{}} = \mathbb{E}(\mu_{i,t}|\theta_{i,t}) = \left[ \frac{\sigma_{\mu_i}^2}{\sigma_{\epsilon}^2 + \sigma_{\mu_i}^2}\theta_{i,t} + \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + t\sigma_{\mu_i}^2}\mu_{i,t-1}^{\hat{}} \right] \quad (4)$$

with variance

$$Var(\mu_{i,t}^{\hat{}}|\mu_{i,t-1}^{\hat{}}, \theta_t) = \frac{\sigma_{\epsilon}^2\sigma_{\mu_i}^4}{(t\sigma_{\mu_i}^2 + \sigma_{\epsilon}^2) [(t-1)\sigma_{\mu_i}^2 + \sigma_{\epsilon}^2]} \quad (5)$$

where, for notational simplicity, we let  $\sigma_{\mu_i}^2$  be the hypothesized variance of the drift parameter from the previous period, or  $\widehat{\sigma_{\mu_i,t-1}^2}$ . The basic intuition is that as more and more signals are received, new signals are downweighed in (4) because they represent a smaller fraction of the overall information available for inference.

In our model, individuals do not replace signals that contradict their hypotheses with signals that confirm their hypotheses. This is important as we assume that individuals correctly distinguish between positive and negative signals. This assumption limits the extent to which individuals “make mistakes.” Individuals do not erroneously infer contradictory evidence as being confirming.<sup>8</sup> While individuals always have confirmatory bias, they replace evidence that *weakly* supports their hypotheses with evidence that *strongly* supports their hypotheses. This allows for belief polarization. Individuals become more and more convinced of their original position as long as there is not significant contradictory evidence, as is commonly found in the psychology literature and the existing models of confirmatory bias. More formally, we define confirmatory bias as follows:

---

<sup>8</sup>As our data is a poll of experts, it is intuitive to assume that they correctly distinguish positive from negative signals.

**Definition 1:** An individual has confirmatory bias if, whenever  $\mu_{i,\hat{t}-1} > 0$ , an observer replaces  $\theta_{i,t} \in [0, \kappa)$  with  $\theta'_{i,t} = \theta_{i,t} + \kappa$  and, whenever  $\mu_{i,\hat{t}-1} < 0$ , an observer replaces  $\theta_{i,t} \in (-\kappa, 0]$  with  $\theta''_{i,t} = \theta_{i,t} - \kappa$ .

Definition 1 operationalizes confirmatory bias by explicitly replacing one signal with a stronger signal. However, the magnitude of the replacement does not depend on the magnitude of the signal; *a priori*, we have no reason to suspect that agents react to certain ambiguous signals more or less strongly than other signals.<sup>9</sup>

Figure 1 shows the regions of confirmatory bias in our model. The figure shows the perceived signal distribution from the perspective of the agent. Although signals are normally distributed in this environment, individuals with confirmatory bias treat the signal distribution as if it were missing a large mass near 0, where the signal indicates that the team is weakly improving or declining in quality.<sup>10</sup> For example, if the initial hypothesis is positive, then as depicted in Figure 1,  $\hat{\mu} > 0$ , and when individuals receive a signal slightly greater than or equal to zero from the interval  $[0, \kappa)$ , they replace that signal with a signal from the interval  $[\kappa, 2\kappa)$ , which strengthens their hypothesis that the drift term is positive and that the team is improving in quality. The results are similar for a negative hypothesis.<sup>11</sup> In particular, in this environment, when an individual has confirmatory bias and a positive hypothesis, their expected estimate of the drift term changes as follows:

$$\mathbb{E}(\mu_{i,t}|\theta_{i,t}) = (1-F(\kappa)+F(0)) \left[ \frac{\sigma_{\mu_i}^2}{\sigma_{\epsilon}^2 + \sigma_{\mu_i}^2} \theta_{i,t} + \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + t\sigma_{\mu_i}^2} \mu_{i,\hat{t}-1} \right] + (F(\kappa)-F(0)) \left[ \frac{\sigma_{\mu_i}^2}{\sigma_{\epsilon}^2 + \sigma_{\mu_i}^2} \theta'_{i,t} + \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + t\sigma_{\mu_i}^2} \mu_{i,\hat{t}-1} \right] \quad (6)$$

It is important to note that this is an expected value for the drift term. The actual estimate for

---

<sup>9</sup>For example, multiplying the signal by a constant would amplify different signals by different amounts such that agents would be reacting more strongly to some ambiguous signals rather than others, and adding a constant that is a function of the signal would cause the magnitude of the bias to be sensitive to a function with unknown properties.

<sup>10</sup>Signals are normally distributed because they are sums of two i.i.d. normal random variables,  $\mu$  and  $\epsilon$ .

<sup>11</sup>Since we use ranking data for the 25 best teams our empirical estimates relate to positive hypotheses.

$\mu_{i,t}$  depends on the realization of  $\theta$ :

$$\hat{\mu}_{i,t} = \begin{cases} \frac{\sigma_{\mu_i}^2}{\sigma_\epsilon^2 + \sigma_{\mu_i}^2} \theta'_{i,t} + \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2 + t\sigma_{\mu_i}^2} \mu_{i,\hat{t}-1} & \text{if } \theta \in [0, \kappa); \\ \frac{\sigma_{\mu_i}^2}{\sigma_\epsilon^2 + \sigma_{\mu_i}^2} \theta_{i,t} + \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2 + t\sigma_{\mu_i}^2} \mu_{i,\hat{t}-1} & \text{if } \theta \notin [0, \kappa). \end{cases} \quad (7)$$

In particular, the limit as the signal approaches zero is:

$$\lim_{\theta \rightarrow 0^+} \mathbb{E}(\mu_{i,t} | \theta_{i,t}) = \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2 + t\sigma_{\mu_i}^2} \mu_{i,\hat{t}-1} + (F(\kappa) - F(0)) \left[ \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_{\mu_i}^2} \kappa - \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_\epsilon^2}{\sigma_\epsilon^2 + t\sigma_{\mu_i}^2} \mu_{i,\hat{t}-1} \right] \quad (8)$$

As the signal approaches zero (indicating that the drift term is neither positive nor negative), there is still bias as described by the first term in the square brackets,  $(F(\kappa) - F(0)) \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + t\sigma_{\mu_i}^2} \kappa$ . This term represents the proportion of the estimate of the drift term that is altered in expectation due to an individual's confirmatory bias. It is increasing in  $\kappa$ , which we interpret as the severity of confirmation bias for an individual. As  $\kappa$  gets large the severity of the bias increases. This is intuitive because the individual makes larger mistakes when he augments the old signal with a larger signal. Furthermore, the probability that the individual will continue to make mistakes going forward increases as well, because the probability of receiving a relatively strong signal expands from  $(F(2\kappa) - F(\kappa))$  to  $(F(2\kappa) - F(0))$ . In future periods, the probability of mistakes increases because beliefs become more polarized; a greater number of negative signals or more intense negative signals are necessary in order to drive the hypothesis in the other direction.<sup>12</sup>

In our model,  $\kappa$ , the magnitude of how an individual over-infers from a weak signal, remains constant over time, and thus the amount of confirmatory bias from any one signal is constant over time. This is clearly seen in (7), where the only differences between announcements come from the altered signal  $\theta'_{i,t}$ , which is a function of the original signal and  $\kappa$ . If an individual receives the same weak signal in different periods, i.e.  $\theta_{i,t} = \theta_{i,-t}$ , then he responds to both  $\theta_{i,t}$  and  $\theta_{i,-t}$  by adding the same amount,  $\kappa$ , to the signal both times. In this way, we only have one source of belief polarization, the accumulation of signals over time. While individuals become more convinced of their hypotheses over time because of the aggregation of altered signals, they do not react in a more

---

<sup>12</sup>Along the same lines, our model also implies that individuals make more mistakes as  $\kappa$  increases, since the range of signals that they misinterpret is larger.

confirmatory way to the same signal as time elapses. We contend that belief polarization occurs gradually; people are more likely to reverse their hypothesis when they have altered signals only a few times and less likely to change their hypotheses after altering signals many times.

Our dynamic model of updating with confirmatory bias yields equation (8), the limit of an individual’s estimate of the mean of the drift term as the signal becomes weak, which motivates the regression-discontinuity design that we exploit below. If an individual did not have confirmatory bias, then (8) would reduce to:

$$\lim_{\theta \rightarrow 0^+} \mathbb{E}(\mu_{i,t} | \theta_{i,t}) = \frac{(t-1)\sigma_{\mu_i}^2 + \sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + t\sigma_{\mu_i}^2} \mu_{i,t-1}. \quad (9)$$

Because the fraction multiplying  $\mu_{i,t-1}$  is less than one, when more and more signals equal or infinitely close to zero are received by the agent, the agent’s beliefs about the mean of the quality distribution converge to 0 as  $t \rightarrow \infty$ . Intuitively, if an agent receives a number of signals that a team is neither improving or declining in quality, then the announcements for  $q_{i,t}$  and  $q_{i,t+1}$  will eventually converge to each other. However, with confirmatory bias this is not the case: under the same scenario beliefs about  $\mu_i$  can converge to *at best*  $\kappa$ . Practically, this means that an agent’s hypothesis is that a team is always improving in quality even when receiving signals that would indicate that quality remains the same.

For identification, we estimate how beliefs change as signals become less consistent with strongly positive quality improvements, and test to see whether or not announcements contain confirmatory bias, expressed as a discontinuity at the point where the signal becomes ambiguous. We argue that the discontinuity between weakly negative and positive signals should trigger confirmatory bias since it represents a focal point. The discontinuity is the point where individuals have the “opportunity” to interpret weakly positive evidence as if it were strong evidence. In this way, our model encapsulates the key insight from other models of confirmatory bias.



## 4 Identifying Confirmatory Bias in the Field

The key advantage of our empirical approach is that we exploit a quasi-experimental setting to measure both the magnitude and prevalence of confirmatory bias, allowing us to distinguish confirmatory bias from both pure Bayesian updating and other psychological and cognitive biases. We use historical data to test for confirmatory bias in a real-world setting, where experts announce their beliefs about quality after observing a discrete number of signals between belief announcements. The Associated Press Top 25 College Football Poll (AP Poll) is well-suited for this purpose. As described in Sinkey (2011), it is unique in that it is relatively free from perverse incentives that cause individuals to shade predictions. For example, this environment is different from stock or commodity analysts, credit rating agencies, or housing price indices such as the Case/Shiller index, in which individuals might actively look to manipulate prices. Individuals in this poll are not paid for their opinions. At a minimum, they have few incentives to announce shaded beliefs for personal or professional gain.

The poll is conducted weekly and only one game is played between polling, usually on Saturday. This feature allows us to observe both prior and posterior beliefs about a team’s quality to construct precise measures for updating. Voters observe how a team performs—whether a team wins or loses, how a team performs relative to market expectations, etc.—and are asked by the Associated Press to rank the 25 best teams for their ballots late Saturday evening. This ranking is aggregated using a Borda count and the teams with the 25 highest aggregate point totals comprise the AP Poll, which is released on Sunday.<sup>13</sup> An important feature of the poll is that individual ballots are made public, which acts as a further deterrent against less-than-truthful (or wildly inconsistent week-to-week) announcements. As we noted earlier, the attention that the public gives to the poll also exerts an influence to provide truthful announcements. Because voters have incentive to be truthful, we are able to use the poll as a measure of internal consistency in information processing, which does not require pollsters’ announcements to be accurate, only consistent.<sup>14</sup>

---

<sup>13</sup>Teams are given points from one to 25 from each individual ballot, with one representing the 25th-best team (the team ranked 25 a pollster’s ballot) and 25 representing the best team (the team ranked one a pollster’s ballot).

<sup>14</sup>Note that, because a team’s true quality can never be perfectly observed, it is impossible to verify whether voter announcements are indeed “accurate.”

Unlike other settings that may also have unbiased experts, this setting only has one “set” of key signals between the belief announcements, since teams only play one game a week. Additionally, this feature differentiates the poll from other possibly useful opinion polls, such as the Survey of Professional Forecasters, where voters observe many different macroeconomic variables, and the AP College Basketball Poll, which is similar to the football poll, but different in that teams may play multiple games in a week. In these alternative settings, the presence of multiple relevant signals between observations poses an identification problem in that one may erroneously attribute a change in beliefs to a bias rather than a reaction to the information provided by a different set of signals.

#### 4.1 Data

We constructed a unique dataset that contains detailed game and betting market results for 31 of the most prominent teams in college football for the years 1985-2008.<sup>15</sup> The teams are listed in Table 1. We note that choosing teams at random would give us relatively few observations as a team must be ranked in the poll. Also, in order to establish exchangeability, we need measures of other observable characteristics not accounted for in the AP poll itself, such as opponent strength and the points scored in individual contests. The teams included in our data are the teams most-consistently ranked (week to week) during this time period. While sample selection is always a concern, an excluded team would have to be one that was consistently ranked *and* played an extraordinarily large number of contests where the margin of victory was close to the betting line for their exclusion to bias the results. We view this possibility as highly unlikely as our data consists of the teams most likely to be ranked over the time period considered.<sup>16</sup> Most important, our data includes a rich set of game characteristics: game location, both the team’s and the opponent’s contemporaneous and season winning percentages, poll rankings of teams and their opponents, and poll points in the AP Poll. The game characteristics were hand-collected from the *ESPN College Football Encyclopedia* (McCambridge (2005)) and were then matched to the detailed historical AP poll data. We include betting lines taken from sports handicapper Jim Feist’s workbook, which

---

<sup>15</sup>See Logan (2011) and Sinkey (2011) for more on sample design.

<sup>16</sup>The majority of teams in the data go several years between polls in which they are not ranked.

provides betting lines for each game over this period.<sup>17</sup> Each observation in our sample consists of a team's performance in a given week of a given season.

The summary statistics for the total sample are listed in Table 2. Teams in our overall sample are more likely to win, as they are stronger teams, and are favored to win by slightly more than eight points, on average. This is consistent with this set of teams being among the most prominent and having long-standing winning traditions. They exceed market expectations roughly half of the time, however, as shown in both the beat spread variable and the difference between the betting line and the margin of victory. In general, their opponents have usually won one more game than they have lost at the time they play the teams in our sample.

Other covariates in the data include the change in the number of AP points after the game, the score of the team, the score of the opponent, the week of the season, the AP rank before the game, the AP rank after the games, indicators for whether the team won or loss the game, and the wins and losses of the opponent. The sample necessarily focuses on ranked teams as these are the only teams for which we have a reliable measure of perceived quality. We note that, as ranked teams, it is likely that voters have positive priors about these teams relative to unranked teams.

The dependent variable is the week-to-week difference between the points a team has in the Associated Press Poll before and after the reference game is played. Since the AP Poll is a Borda count of individual pollster ballots, it is a measure of the change in the perceived quality of a team from individuals who are "experts" on college football and designed to achieve a consensus ranking of teams from a large number of experts. Since the poll points generate the ranking they are a sufficient statistic for the ranking. Another advantage of poll points is their ease of interpretation. Since the number of voters is known, the change in points can be interpreted as the number of pollsters who change their ranking. For example, a positive difference of one point is equivalent to one pollster ranking a team one slot higher than in the previous week.

Our running variable is the difference between the actual margin of victory for a contest and

---

<sup>17</sup>Betting lines are released at the start of the business week, primarily Monday or Tuesday, and are updated up until the start time of the game. The last betting line is known as the "final line." In our estimates, we only use final lines, as this represents the full set of information available right up to the start of the game. Betting lines have been shown to be extremely accurate predictors of actual margins of victory. For example, (Logan and Sinkey, 2011) use all Division 1-A football outcomes from 1985 to 2005 and regress the actual margin of victory on the betting line and find the constant to be -0.00866 (0.11 s.e.) and coefficient on the betting line of 1.003 (0.01 s.e.).

the point spread (the predicted margin of victory) in the betting markets for the same contest. We conjecture that the quantity of new information contained in the signal varies with respect to the magnitude of the difference. Large differences between the actual margin of victory and the point spread likely signal “new” information, as a large divergence between the realized performance and expected performance indicates that the expected performance was particularly inaccurate. In these cases, we would expect observers to update their beliefs significantly: large departures from expectations imply that previous beliefs were inaccurate. We note that large differences could be due to inaccurate priors about quality or random error, but they are observationally equivalent and could be used by experts as the basis of a revised prior.

Small differences between realizations and market expectations, however, provide scant “new” information, as a small difference indicates that the team performed as well as expected. If agents update their expectations using Bayes’ rule, then a small difference between the actual margin of victory and the predicted margin of victory will result in small changes in the perceived quality of a given team. More importantly, these small updates would be symmetric around the point where the actual and predicted margin of victory were equal, small departures in either direction would be weighed equally. If agents drastically alter their perception of a team’s quality when confronted with small departures from expectations, however, then we interpret this as evidence of confirmation bias.

## 5 Methodology

In the majority of studies that employ the regression discontinuity (RD) design, the focus is on estimating the effects of treatment on the outcome of interest where treatment is determined by the value of the running variable. Typical application of the regression discontinuity design focuses on a narrow window around the cutoff because the validity of the causal inference depends on the following assumption: agents who barely fail to qualify for the treatment and those who narrowly qualify are exchangeable. If this assumption holds, then researchers can recover an unbiased estimate of the treatment effect for the subset of the population “close” to the threshold. The RD design appeals to researchers because it offers a means to provide evidence that supports

the assumptions necessary for valid causal inference. In particular, it exploits the insight that the outcome of interest may well be affected by the running variable, but the relationship between the running variable and the outcome variable is smooth, not discontinuous. As long as individuals are unable to precisely control their treatment, the RD approach allows for discontinuities at the cutoff to plausibly be interpreted as causal effects of treatments (Lee, 2008).

We employ the RD design in a somewhat novel, but still appropriate, context that is supported by the underlying behavioral model. Many papers on Bayesian updating need to estimate a belief structure and compare actual responses to this belief structure, which would rely on imposing structural conditions on the exact form of individual belief functions. We are able to avoid this potential drawback by exploiting the fact that voters have no ability to manipulate the running variable, because they neither participate in games nor set betting lines. Additionally, if the assumption of exchangeability holds voters would not update their beliefs in response to other information as it is indistinguishable on either side of the discontinuity. Nevertheless, we allow for the functional form for voter beliefs about quality to differ on each side of the discontinuity, following the recommendation in Lee and Lemieux (2010) and implementing the RD design by estimating:

$$\Delta q_i = \alpha + \tau D + \sum_{k=1}^K \beta_k \theta_i^k + \sum_{k=1}^K \delta_k (D \times \theta_i^k) + \varepsilon_i \quad (10)$$

$D$  is an indicator that assumes a value of one if the difference between the actual margin of victory and the point spread exceeds zero. The running variable,  $\theta$ , is the difference between the actual margin of victory and the expected margin of victory. Like Matsudaira (2008) and Dell (2010), we also estimate specifications with low order polynomials that are fully interacted with  $D$ . This allows the expectation of  $\Delta q_i$  to exhibit different behavior on both sides of the discontinuity. We cluster the standard errors on different values of  $\theta$  due to uncertainty in the choice of functional form for RD designs with discrete support (see Lee and Card, 2008, among others).

Using the difference between the actual margin of victory and the point spread as our running variable eliminates the problems associated with “heaping,” as described in Barreca, Lindo, and Waddell (2011) and manipulation of the running variable as described in McCrary (2008). Heaping

in the running variable may cause RD designs to have biased estimates as measurement error in the running variable can lead to misassignment on either side of the discontinuity. We do not have measurement error in either the betting line or the actual margin of victory. Our data are neither rounded nor self-reported, which are two common causes for heaping. More importantly, we have good reason to think that our data would not heap at any particular difference between the point spread and margin of victory, since if this were known to bettors, they would strategically alter their betting behavior to try and make profit, and, knowing this, betting houses would adjust their lines. Logan and Sinkey (2011) provide evidence that betting houses are good at doing this. McCrary (2008) devises a test to detect manipulation of the running variable; however, as stated earlier, given that voters are able to manipulate neither the betting line, final score, nor the margin of victory in a given contest and that the observations are comparably distributed across the threshold, we assert that there is a clear lack of manipulation.

The departure from Bayesian updating,  $\tau$ , is the parameter of interest. In the model, it is given by  $\frac{\sigma^2}{\sigma_\epsilon^2 + \sigma^2} \kappa$ , which is the amount of confirmatory bias added onto the original estimate of  $\mu_{t,i}$  in any period. If the AP pollsters update in a Bayesian fashion, then the function that relates perceived quality to the quantity of underlying new information will trend smoothly through the point where teams just miss or just exceed expectations. When the difference between the actual margin of victory and the spread approaches zero, little new information is revealed; by definition, the team performed nearly as well as expected in a given contest. Therefore, our model predicts that at points where there is minimal information, there should be no discontinuous adjustments to perceived quality. If  $\tau$  is indistinguishable from zero, then AP pollsters announce their estimates of perceived quality in a manner consistent with Bayesian updating without confirmatory bias. If  $\tau$  assumes a positive value that is statistically distinguishable from zero, then our model characterizes this as evidence that AP pollsters do not update using Bayes' rule and instead exhibit confirmatory bias. Our estimate of  $\tau$  is the local average treatment effect of exposing voter beliefs to an ambiguous signal, and we interpret positive values as strong evidence of confirmatory bias.

## 6 Empirical Results

### 6.1 Establishing Exchangeability

If the baseline characteristics of the team trend smoothly through the discontinuity, then this is evidence that the teams within the window are exchangeable. Such evidence supports the assumption that any observed change in the allocation of poll points is due to over-reaction to the weak signal of barely exceeding the spread and not the provision of new information or other features of the updating process. Because there are no differences in observables between the observations on either side of the discontinuity, the belief changes we observe are caused by the treatment of beliefs with ambiguous signals, which allows us to plausibly interpret our estimate of confirmatory bias as a local average treatment effect. Theoretically, we do not need to model the structure of the updating function; because other signals are equivalent on both sides of the discontinuity, there is no plausible updating due to these signals. Substantively, other signals contain information related to quality, but these signals are equivalent on both sides of the discontinuity. The only variation we have is in barely missing or barely exceeding the spread, which contains very little information about quality since margins of victory in football are incredibly difficult to manipulate in small amounts by design.

To test for this smooth trend through the discontinuity, we estimate a system of Seemingly Unrelated Regressions (SUR) to determine if the baseline characteristics are locally balanced (see, for example, Zellner, 1962). Specifically, we estimate:

$$\begin{aligned} C_1 &= \eta_1 + D\beta_1 + \epsilon_1 \\ &\vdots \\ C_J &= \eta_J + D\beta_J + \epsilon_J \end{aligned} \tag{11}$$

The  $C_j$  are a series of baseline characteristics that may affect updating, including whether a team wins, the opponent's record, and the scores of the team and opponent.  $D$  is a dummy variable

that assumes the value one if the difference between the actual margin of victory and the point spread is larger than zero. Differences across the threshold in the baseline characteristic will be captured by  $\beta_j$ . The SUR approach allows for correct inference even if the  $\epsilon_j$  are correlated across equations. In particular, it accounts for the possibility that, due to the large number of covariates, some discontinuities may occur by random chance. We test to see if each  $\beta_j$  is statistically different from zero.

We choose a window where the difference between expected and actual outcomes is less than or equal to three and one-half points (in absolute value). We choose this window because 3 points is the modal minimum scored in football contest and to account for the fact that betting lines are sometimes set at half-point intervals.<sup>18</sup> Since we concentrate on a very small window where the betting line accurately predicts the margin of victory, our sample consists of 417 contests with a difference of less than three and one-half points between the actual margin of victory.<sup>19</sup> Table 3 shows the balance of the covariates in our sample both above and below the discontinuity. Columns 1-3 of Table 3 indicate that the distributions of the covariates appear to be balanced both above and below the threshold.<sup>20</sup> The third column contains the results of a two-sided t-test comparing the difference in means for the populations above and below the threshold. We find no evidence that the plausibly important factors—for example, the number of points the opponent scores and whether the team won the game—are statistically different for teams that are just above and just below the threshold. The results from column 4 show that all of the estimates in the SUR system are statistically indistinguishable from zero. Taken together, columns 3 and 4 provide strong evidence for balance on both sides of the discontinuity. These results corroborate the assumption that teams just above and just below the threshold are exchangeable, that just exceeding market expectations is uncorrelated with other features of the game in question.

---

<sup>18</sup>Teams can only score one point *after* scoring a touchdown, which is worth six points. Two points can only be scored either after a touchdown or as a result of a safety (which requires an offense to be tackled in their own endzone), which is rare in football.

<sup>19</sup>Note that our identification strategy uses roughly 5% of the total sample.

<sup>20</sup>Since teams on both sides of the discontinuity win their games more than 80 percent of the time our results are not due to double-counting the same contest between two ranked teams.



## 6.2 Graphical Evidence

The motivation for our RD design is visually evident. In Figure 2 we plot the week-to-week changes in poll points (y-axis) and the difference between the actual and expected margins of victory (x-axis) for the entire sample. The figure shows that teams which perform extremely poorly, with differences substantially less than zero, are thought to be of much lower quality than expected, as they have large negative poll point changes, on average. Teams which perform extremely well relative to expectations are thought to be of much higher quality than expected, as they have large positive poll point changes. At zero, however, there is a visible discontinuity in belief changes. Teams whose performance is just to the right of zero (who slightly exceed expectations) are assessed as being of significantly higher quality than teams whose performance is just to the left of zero. Indeed, the slopes of the linear updating function differ on either side of the discontinuity.

This discontinuity persists despite the fact that, in many instances, betting lines are set in half-point increments, and thus a team’s performance is *automatically* forced to one side or the other of the discontinuity. In these games, it is impossible for poll voters to set their expectations that a team will win or lose by a half-point increment because teams are unable to score half-points. In these types of scenarios, which happen frequently in the betting market, it would be especially strange to value close differences between the margin of victory and the betting line since those differences have to occur by default. Indeed, half of our support in the running variable, irrespective of the window size, consists of such games. (As a robustness check, we also show results for other “windows”—two and one-half points, three points, four points, and four and one-half points—in the next section.)

## 6.3 Regression Discontinuity Estimates of Confirmatory Bias

Table 4 contains the results from the RD design for all specifications for our window of three and one-half points. The standard errors are clustered on the running variable,  $\theta$ , which allows us to account for potential misspecification of the conditional mean function. Column 1 of Table 4 contains the results of a basic linear regression in equation (10) of the difference in AP Poll points from before and after the contest on an indicator variable that assumes a value of one if the team

beats the spread in a given contest. The basic regression indicates that barely exceeding the point spread results in a 29.74 point gain in AP Poll points. Put another way, the results are equivalent to nearly half of the AP pollsters rank a team one slot higher for barely beating the spread, as there are 60 voters in the AP poll. This estimate is statistically different from zero. Its magnitude is considerable, as this gain is nearly 19 percent of the standard deviation of the change in AP Poll points for the sample.<sup>21</sup> Furthermore, we find that this result is robust to the selection of different windows. When we change our windows to two and one-half, three, four, and four and one-half point windows (shown later), we find that beating the spread on voter beliefs continues to be significant for all our basic specifications.

Columns 2-4 in Table 4 contains the estimate of the discontinuity with a specification that includes fully interacted quadratic, cubic, and quartic polynomials in  $\theta$ , respectively. These estimates allow the conditional expectations function to differ on both sides of the threshold with increasing flexibility, and are designed to account for potential misspecification of the belief function on either side of the threshold. Even with this additional flexibility in the conditional expectation function, we find significant evidence that voters take into account beating the spread when updating their beliefs. In the quadratic specification, voters add 81.23 points to their belief function and that this result is significant at the ten percent level. In our cubic and quartic specifications we find the estimated effect is quite large, at 210.03 points, which is 133 percent of the standard deviation, and 308.53 points for the quartic, which is a staggering 196 percent of the standard deviation.

The point estimates in Columns 2-4 are quite large. These estimates likely exacerbate the true response, as estimating higher-order models in small bandwidths overfit the data (see Lee and Lemieux (2010)). When we condition our estimates of  $\tau$  on the expected values of  $\theta$  for the flexible functional forms employed, however, they equal the results in Column 1. As such, these results lend credence to our assertion that voters use imprecise information as confirmatory when they already have positive hypotheses of team quality.

Figure 3 provides graphical evidence of our results. In this figure, each point represents the

---

<sup>21</sup>The standard deviation of the change in AP Poll points for the estimating sample is 157.26 points. Similarly, Sinkey (2011) finds that total updating for beating the spread is anywhere from 50 to 70 points depending on the specification.

average of the change in AP Poll points for each half-point interval, and each curve represents a separate regression of the means of the differences in AP Poll points on quadratic, cubic, and quartic polynomials in  $\theta$ , our running variable. Each curve provides clear visual evidence of a discontinuity in beliefs at zero. In addition, the polynomials in this figure show how our belief estimates from the conditional mean function evolve as additional flexibility is added to the functional form. This provides strong graphical evidence of a large jump in beliefs at the threshold.

We interpret these findings as evidence of confirmatory bias. Votes in the poll necessarily represent positive hypotheses about quality of teams, because these teams are selected as being among the 25 best teams in the country.<sup>22</sup> In our model, when voters with positive hypotheses are confronted with weakly positive signals about team qualities, they replace these signals with more positive signals, generating persistence in confirmatory beliefs over time. The results are consistent with the model.

## 7 Extensions and Robustness

### 7.1 Extensions: Constant Bias Over Time

In our model, it becomes increasingly difficult to convince an individual to change his or her hypothesis if they continue to receive weak signals. This leads to an unwillingness to change one's hypothesis in response to new data, which may lead to sub-optimal decision making. Our model allows us to examine the source of belief polarization. In particular, we concentrate on equation (7), which indicates that the magnitude of confirmatory bias in response to any one signal should not increase as signals accumulate. In (7), individuals replace weak signals in the region  $\theta_{i,t} \in [0, \kappa]$  with new signals  $\theta'_{i,t} = \kappa + \theta_{i,t}$ . An important distinction of our model is that the parameter  $\kappa$  remains fixed for any period  $t$ ; it is time-invariant. Belief polarization occurs because individuals have accumulated many augmented signals over time. While the cumulative effects may be large, the bias at any point is the same. This is intuitive within the framework of confirmatory bias—individuals do not know that they have confirmatory bias, and as such their over-inference is a

---

<sup>22</sup>Teams in the poll are thought to be among the top twenty-five percent of teams in Division 1-A.

constant (time invariant) cognitive bias.

If the amount that individuals erroneously infer remains constant over time, then we should not expect to find significantly different responses to weak signals at different parts of the season since we are looking at differences from week to week and are thus conditioning on prior estimates of quality. As the season progresses (the week of the season increases), individuals receive more signals about a team’s quality, and beliefs may become more polarized. However, our model stipulates that belief polarization does not imply that individuals are reacting more or less to weak signals, but rather their cumulative effects.

It is also important to note that the pollsters in our setting are experts. It is natural to assume that beliefs early in the season would be based on scant evidence, as the season progresses one would expect experts to exhibit less bias over time. In particular, the *within* season experience of the pollsters implies that biases such as over- or under-reactions would diminish over time. Intuitively, we would expect experts to learn quickly over the course of a season. As such, we view this extension as a powerful test of the cognitive nature of confirmatory bias.

We test to see whether there are any differences in updating that depend on the week of the season by estimating (10) with the inclusion of a variable that captures the week of season interacted with beating the spread. Formally, we estimate:

$$\Delta q_i = \alpha' + \tau' D + \gamma D * WOS + \sum_{k=1}^K \beta'_k \theta_i^{k'} + \sum_{k=1}^K \delta'_k (D \times \theta_i^{k'}) + \varepsilon_i' \quad (12)$$

where *WOS* represents the week of the season and  $\gamma$  represents the magnitude of how individuals respond to beating the spread in a particular week of the season. If individuals respond to weak signals more significantly later in the season, then we would expect  $\gamma$  to be positive and significant. However, if individuals do not respond to weak signals more significantly later in the season, then we should not expect this interaction term to be statistically different from zero.

Table 5 shows the results. Column 1 reports our results of a basic linear regression of the difference in AP Poll points from before and after the contest on an indicator variable for beating the spread and the interaction variable designed to capture whether individuals respond more to weak signals as the season progresses. While our estimate for beating the spread remains similar to

our estimates in Table 4, we do not find any evidence that individuals react more to weak signals later in the season. Columns 2-4 corroborate the results from our basic specification. As we allow the conditional expectations function to vary, we again find the voters react to weak signals, but do not react more strongly later in the season compared to earlier in the season. The results of Table 5 are consistent with the model.<sup>23</sup>

These results provide evidence for the mechanics of belief polarization, and also serve as a check against alternative explanations. If individuals more strongly augmented their beliefs over time based on when they received information, then this augmentation should be captured in our estimates, since these estimates represent the marginal upgrade to individual beliefs based on when information is received. Because we control for when weak information is received, we are able to disentangle different explanations for why individual beliefs become more polarized over time.

As an example, our results are inconsistent with the idea that voters may learn to de-bias their opinions over time. A common criticism of laboratory studies of cognitive biases is that, given sufficient experience, individual biases would be reduced or eliminated. In our setting, if this criticism were true, subsequent observations would temper confirmatory bias. However, individuals do not “improve” their signal processing with time, as beating the spread remains significant throughout the season. Similarly, our results are inconsistent with the idea that belief polarization occurs because people more strongly react to information at later dates. It is more likely that polarization occurs because of many separate overreactions to weak signals, which, in our model, occurs due to confirmatory bias.

## 7.2 Robustness Checks: Sensitivity to Bandwidth Selection

As a sensitivity check, we test to see whether our estimates for the magnitude of confirmatory bias are sensitive to bandwidth selection. While we use three and one-half points as our base estimation window, we also provide estimates from both larger and smaller windows. Windows of two and one-half, three, four, and four and one-half points are arguably identical in information content

---

<sup>23</sup>We estimated models with squared and cubic terms for week of the season for greater flexibility in the week of season effect and those results were not statistically significant. We also estimated versions of equation (12) with dichotomous indicators for thirds of the season and the results were not statistically significant in those specifications, either.

to voters. Since these windows also provide scant new information about a team’s quality, we should expect that voters with confirmatory bias will react similarly to beating the spread in these windows as well. We note, however, that as the window expands our model predicts that some of the resulting updating may be completely rational since the signal is becoming more informative the larger the departures are from expectations.<sup>24</sup> (We provide statistical evidence for balance in these windows and graphical evidence for the discontinuity the appendix.)

Table 6 contains the results of our RD design for the two and one-half, three, four, and four and one-half point windows, respectively. Column 1 provides estimates from a basic linear regression of the change in AP Poll points on an indicator variable that assumes a value of one if a team beats the spread in a given contest. In the three, four, and four and one-half point windows, beating the spread is a significant determinant of realized belief changes.<sup>25</sup> This evidence is quantitatively similar to our results from our main bandwidth. We find that beating the spread adds 29.11 points to a team’s ranking in the three point window, 33.96 points to a team’s ranking in the four point window, and 42.53 points in the four and one-half point window. Practically, this is equivalent to between half to three-quarters of voters moving a team up one slot in the poll, depending on the window. Consistent with our earlier argument, the size of the change in belief increases with the size of the window by a small amount. Columns 2-4 of Table 6 provide evidence that our result is robust to inclusion of higher order terms in the running variable.

Beating the spread is a highly significant determinant of voter beliefs. We find compelling evidence that beating the spread is significant for a wide range of specifications and is robust to potential misspecification of the belief function.<sup>26</sup> These estimates provide robust evidence for confirmatory bias in our sample. Pollsters seem to over-infer from weak information across both bandwidths and specifications.<sup>27</sup>

---

<sup>24</sup>We note that the graphical evidence in Figure 2 shows that the changes in poll points increases slowly as the departure from expectations increases.

<sup>25</sup>Beating the spread is also marginally significant in the two and one-half point window, just outside the ten percent level.

<sup>26</sup>As we add fully-interacted polynomials in  $\theta$ , we again find that our point estimates are quite large and may exacerbate the true response, as described in (Imbens and Lemieux (2008) and Lee and Lemieux (2010))

<sup>27</sup>In the appendix we estimate a fully interacted specification which includes additional controls that have been shown to be related to voter updating (beating the spread in the previous week, winning the game, etc.). The results from these specifications are qualitatively similar to the results reported here.

## 8 Conclusion

Cognitive biases may have significant impacts on realized economic decisions, but evidence of the extent to which these biases affect decisions is sparse for many reasons. The very nature of the market may mitigate the effects of individual bias in decision-making. Similarly, more experienced agents are less likely to exhibit cognitive biases. Empirically, it is difficult to both match reliable estimates of individual beliefs to the information used to construct those beliefs and to track beliefs and information over time. Because of these limitations, learning about either the magnitude or persistence of a bias in a real-world setting is difficult.

We are able to track hypotheses over time and are able to estimate how individuals respond to new information. We utilize a regression discontinuity design to test for the magnitude of confirmatory bias in the field, allowing us to focus on the response to a specific signal. We focus on the unique variation afforded by examining how beliefs change when confronted with small differences between realized and expected results. We find that individuals react to small positive differences differently from small negative differences. These reactions are inconsistent with Bayesian updating. Our estimates of how individuals respond above and below the threshold are economically large and provide direct empirical evidence of confirmatory bias in a real world setting.

We find that our results are robust to a large number of potential confounds. For example, we find robust evidence to our basic estimates in our quadratic, cubic, and quartic specifications which add increasing flexibility to the conditional expectation function. In each of these specifications, voters substantially value exceeding expectations by small amounts. Similarly, we find that our results are not sensitive to our choice of estimation window. In each of the windows we choose, we find that voters significantly react to small, positive deviations from expectations. We also find that the amount of confirmatory bias that we estimate is constant over time.

Even in an environment where the information available between announcements is very tractable, we find evidence from announced beliefs that individuals may over-infer from weak signals when they are consistent with previous hypotheses. It is reasonable to think that individuals may be subject to similar types of biases in environments where there are many pieces of information to consider and where hypotheses have stronger consequences. These estimates indicate that even

experts exhibit the tendency to drastically alter their perceptions of quality when confronted with weak signals that are largely bereft of new information.



## References

- Al-Ubaydli, O. and J. A. List (2012). On the Generalizability of Experimental Results in Economics. NBER Working Paper No. 17957.
- Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *Quarterly Journal of Economics* 107, 797–817.
- Barreca, A. I., J. M. Lindo, and G. R. Waddell (2011, September). Heaping-Induced Bias in Regression-Discontinuity Designs. Working Paper 17408, National Bureau of Economic Research.
- Card, D. and G. Dahl (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *Quarterly Journal of Economics* 126(1), 103–143.
- Card, D., S. DellaVigna, and U. Malmendier (2011). The Role of Theory in Field Experiments. *Journal of Economic Perspectives* 25(3), 39–62.
- Coleman, B. J., A. Gallo, P. Mason, and J. W. Steagall (2010). Voter Bias in the Associated Press Poll. *Journal of Sports Economics* 11(4), 397–417.
- Darley, J. and P. Gross (1983). A Hypothesis-Confirming Bias in Labeling Effects. *Journal of Personality and Social Psychology* 44, 20–33.
- Dell, M. (2010). The Persistent Effects of Peru’s Mining Meta. *Econometrica* 78(6), 1863–1903.
- Gennaloli, N. and A. Shleifer (2010). What Comes to Mind. *Quarterly Journal of Economics* 125(4), 1399–1433.
- Gottlieb, R. (2010). Will You Never Learn? Self-Deception and Biases in Information Processing. Unpublished Manuscript, Princeton University.
- Imbens, G. and T. Lemieux (2008). Regression-Discontinuity Designs: A Guide to Practice. *Journal of Econometrics* 142(2), 615–635.
- Jones, M. and R. Sudgen (2001). Positive Confirmation Bias in the Acquisition of Information. *Theory and Decision* 50, 59–99.

- Kotchen, M. and M. Potoski (2011). Conflicts of Interest Distort Public Evaluations: Evidence from the Top 25 Ballots of NCAA Football Coaches. NBER Working Paper No. 17628.
- Lee, D. and D. Card (2008). Regression Discontinuity Inference with Specification Error. *Journal of Econometrics* 142(2), 655–674.
- Lee, D. and T. Lemieux (2010). Regression-Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2), 281–355.
- Lee, D. S. (2008). Randomized Experiments from Non-Random Selection in U.S. House Elections. *Journal of Econometrics* 142(2), 675 – 697.
- Levitt, S. and J. List (2008). Homo Economicus Evolves. *Science* 319, 909–910.
- Logan, T. D. (2011). Econometric Tests of American College Football’s Conventional Wisdom. *Applied Economics* 43, 2493–2518.
- Logan, T. D. and M. Sinkey (2011). Does The Hot Hand Drive The Market? Evidence from Betting Markets. Unpublished Manuscript, The Ohio State University.
- Lord, C., L. Ross, and M. Lepper (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology* 37, 2098–2109.
- Mahoney, M. (1977). Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review Process. *Cognitive Therapy and Research* 1, 161–175.
- Matsudaira, J. (2008). Mandatory Summer School and Student Achievement. *Journal of Econometrics* 142, 829–850.
- McCambridge, M. (Ed.) (2005). *ESPN College Football Encyclopedia: The Complete History of the Game*. New York: ESPN Books.
- McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics* 142(2), 698–714.

- Meissner, C. and S. Kassin (2002). “He’s Guilty”: Investigator Bias in Judgments of Truth and Deception. *Law and Human Behavior* 26(5), 469–480.
- Nickerson, R. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2(2), 175–220.
- Rabin, M. and J. Schrag (1999). First Impressions Matter: A Model of Confirmatory Bias. *Quarterly Journal of Economics* 114(1), 37–82.
- Sinkey, M. (2011). How Do Experts Use Bayes’ Rule? Lessons From an Incentive-Free Environment. Unpublished Manuscript, The Ohio State University.
- Stone, D. (2011). Testing Bayesian Updating With the AP Top 25. forthcoming, *Economic Inquiry*.
- Stone, D. and B. Zafar (2010). Bayesian Social Learning, Conformity, and Stubbornness. Unpublished Manuscript, Oregon State University and New York Fed.
- Wilson, A. (2003). Bounded Memory and Biases in Information Processing. Unpublished Manuscript, New York University.
- Yariv, L. (2005). I’ll See It When I Believe It- A Simple Model of Cognitive Consistency. Unpublished Manuscript, California Institute of Technology.
- Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 57(298), 348–368.

Table 1: Teams Used in Sample

Alabama	Georgia	Penn State
Arkansas	Iowa	Southern California (USC)
Auburn	Louisiana State (LSU)	Stanford
Boston College	Miami, Florida	Tennessee
Brigham Young (BYU)	Michigan	Texas
California	Nebraska	Texas A & M
California, Los Angeles (UCLA)	Notre Dame	Virginia Tech
Clemson	Ohio State	Washington
Colorado	Oklahoma	West Virginia
Florida	Oregon	Wisconsin
Florida State		

Table 2: Summary Statistics for Entire Sample

Variable	N	Mean	Standard Deviation	Min	Max
Margin of Victory	8169	8.339	20.933	-77	81
Betting Line	8169	7.462	13.925	-45	57
Beat Spread	8169	.517	.491	0	1
Win	8169	.660	.473	0	1
Opp. Wins-Losses	8169	1.069	3.254	-11	12
Betting Line - Margin of Victory	8169	.876	16.066	-80	94
AP Poll Points Change	4297	-3.387	221.388	-1007	910
AP Rank Change	4363	-.100	3.561	-17	12

Opponent W-L is defined as the number of wins minus the number of losses that a team's opponent has at the time of the game.

Table 3: Balance of Covariates: Within Three and One-Half Points

Variable	Below Zero	Above Zero	Difference	SUR Estimate
	(1)	(2)	(3)	(4)
Week of Season	8.240 (3.115)	8.249 (3.138)	.009 [-.004]	-.781 (2.592)
Score of Team	28.804 (10.769)	30.873 (12.126)	2.069 [1.488]	-3.892 (9.591)
Score of Opponent	17.049 (9.435)	16.911 (9.154)	-.138 [-.144]	1.514 (7.723)
AP Rank Before Game	10.147 (6.292)	10.939 (6.765)	.792 [.959]	4.548 (5.462)
AP Rank After Game	10.196 (6.487)	10.610 (6.734)	.414 [.799]	3.475 (5.524)
Wins of Opponent (By Game)	3.647 (2.416)	3.530 (2.492)	-.117 [-.0495]	-1.245 (2.041)
Losses of Opponent (By Game)	2.412 (1.914)	2.559 (2.022)	.147 [1.007]	.557 (1.644)
Win	.843 (.364)	.883 (.323)	.04 [1.570]	.221 (.287)
Home	.603 (.490)	.521 (.501)	-.082* [-1.69]	-.587 (.414)
<i>N</i>	204	213	417	417

The first and second columns contain means with standard deviations below. The third column contains t-statistics from a Welch's t-test of a difference in means in the covariates, since the two samples have unequal variances. The fourth column contains the coefficient estimates from a system of seemingly unrelated regressions, where each equation in the system consists of a regression of the variable in the first column on an indicator that assumes a value of one if the team beat the spread and the interaction terms in the running variable. In this column, we test to see if the coefficient on the indicator is zero, and report standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 4: Regression Discontinuity Estimates of Confirmatory Bias

	(1)	(2)	(3)	(4)
Beating the Spread	29.736**	81.230*	210.033***	308.533***
	[11.347]	[41.801]	[50.553]	[75.889]
N	417	417	417	417
R-squared	.009	.016	.020	.021
Quadratic in $\theta$ ?		X	X	X
Cubic in $\theta$ ?			X	X
Quartic in $\theta$ ?				X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. Specifications (2)-(4) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .

Table 5: Testing for Constant Confirmatory Bias over Time

	(1)	(2)	(3)	(4)
Beating the Spread	35.413*	84.639*	213.956 ***	310.085 ***
	[16.478]	[41.160]	[50.802]	[75.865]
Beating the Spread*Week of Season	-.688	-.472	-.522	-.448
	[1.551]	[1.723]	[1.735]	[1.763]
N	417	417	417	417
R-squared	.009	.016	.020	.021
Quadratic in $\theta$ ?		X	X	X
Cubic in $\theta$ ?			X	X
Quartic in $\theta$ ?				X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. Specifications (2)-(4) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .

Table 6: Regression Discontinuity Estimates of Confirmatory Bias:  
Sensitivity Check

	(1)	(2)	(3)	(4)
Panel 1: RD Implementation, Two and One-Half Point Window				
Beating the Spread	24.532 [13.787]	145.170*** [44.530]	277.468*** [74.022]	1044.259*** [0.001]
N	284	284	284	284
R-squared	.006	.013	.020	.023
Panel 2: RD Implementation, Three Point Window				
Beating the Spread	29.093** [12.274]	108.259** [38.782]	236.024 *** [68.367]	445.646 *** [125.400]
N	381	381	381	381
R-squared	.008	.017	.019	.020
Panel 3: RD Implementation, Four Point Window				
Beating the Spread	33.949*** [10.218]	61.717 [42.861]	160.937*** [45.433]	308.573*** [66.927]
N	508	508	508	508
R-squared	.012	.016	.022	.023
Panel 4: RD Implementation, Four and One-Half Point Window				
Beating the Spread	42.534*** [11.586]	80.480** [53.382]	92.509* [65.896]	312.992*** [66.927]
N	549	549	549	549
R-squared	.019	.030	.031	.037
Quadratic in $\theta$ ?		X	X	X
Cubic in $\theta$ ?			X	X
Quartic in $\theta$ ?				X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. Specifications (2)-(4) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .

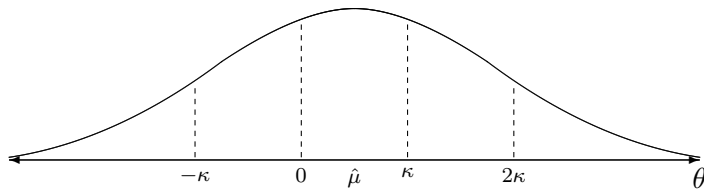


Figure 1: Regions of Confirmatory Bias: *Perceived* Signal Distribution



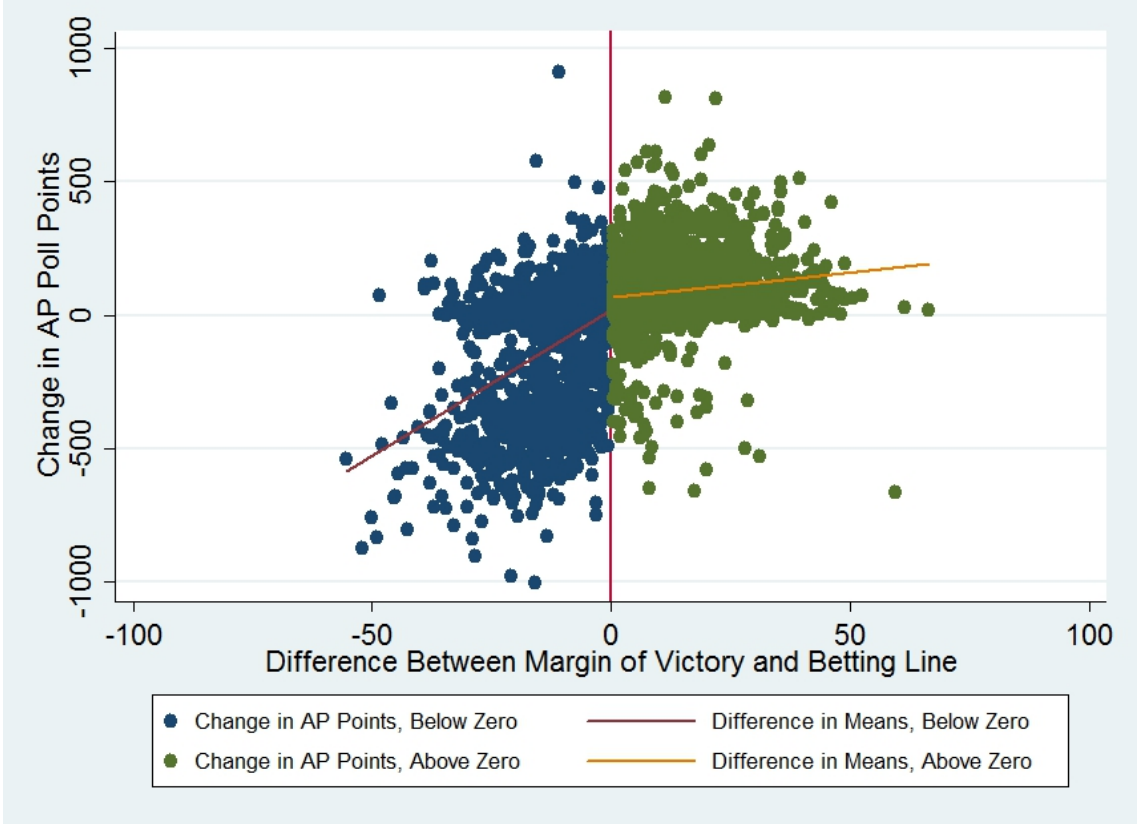


Figure 2: Belief Updating Relative to Expectations, Full Sample

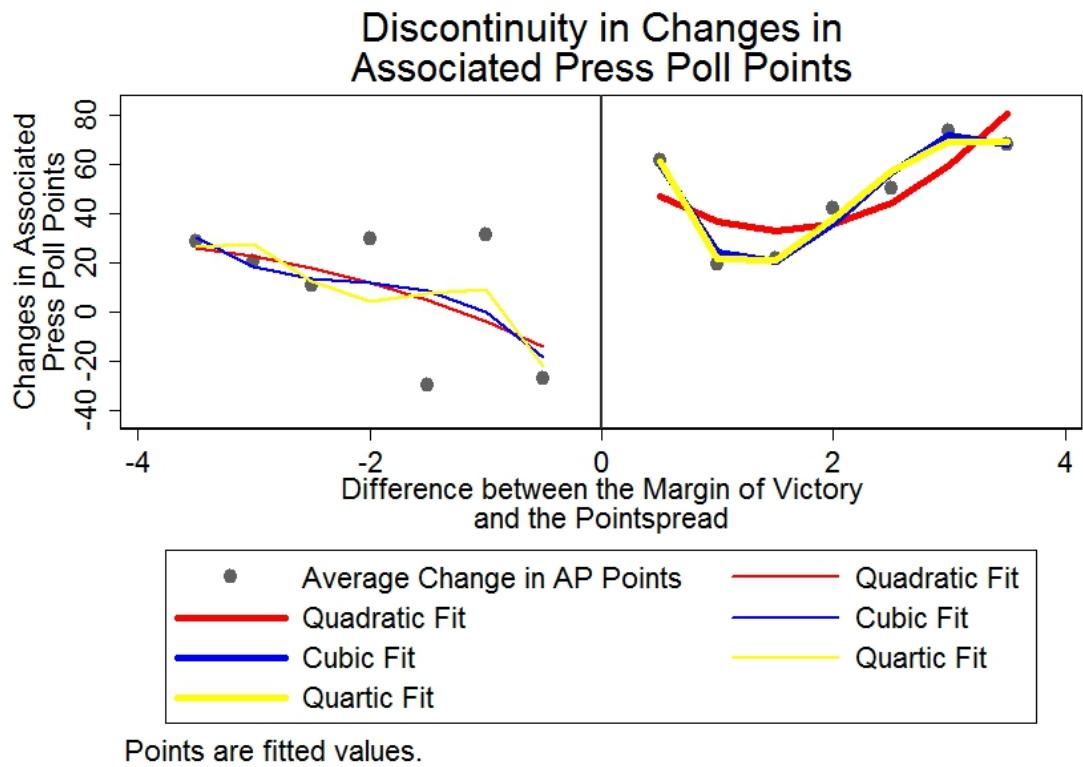


Figure 3: Belief Dynamics at the Discontinuity, Higher Order Polynomial Fits, 3.5 Point Window

# Appendix

## A Evidence of Balance For Different Bandwidths

We show evidence of balance on both sides of the discontinuity for the two and one-half point, three point, four point, and four and one-half point windows, respectively. Appendix Table 1 provides our estimates of balance for the two and one half point window. In this table, we find evidence of balance for a majority of our variables. From our Welch’s t-test, shown in Column 3, we find evidence of balance for all covariates except for the number of points scored by a team, which might be expected since beating the spread necessarily requires a higher score. However, we find no evidence from our SUR estimates that the points of the team is not balanced on both sides of the discontinuity. Only one estimate, the coefficient on opponent losses at time of game, is significant, and only at the ten percent level. Moreover, there is no statistical difference above and below the discontinuity for this variable from our t-test. All of the other coefficients are insignificant. Appendix Table 2 shows evidence of balance for all covariates in the three point window from both a Welch’s t-test, shown in Column 3, and our SUR estimates, contained in Column 4. Again only one measure—the wins of the opponent—is significant and only at the ten percent level. We do not find any statistically significant difference in the means of the covariates for any of the other measures which indicates that there is balance within this window.

Appendix Tables 3 and 4 provide our estimates of balance for the four point window and four and one-half point windows, respectively. According to the t-test in Column 3 of both tables, only the means of three of our covariates, the location of the game, the score of the team and winning, are statistically different from each other above and below the threshold. We might expect these differences because beating the spread necessarily requires a higher score, and teams with higher scores are more likely to win. We do not find evidence that the score of the team and winning are different from our SUR estimates in Column 4, and we only find evidence at the ten percent level that there is a difference in location in the four and one-half point window. Overall, Appendix Tables 1-4 provide strong evidence for balance on both sides of the discontinuity.

Appendix Figures 1-8 replicate Figure 3 and 4 for all of the other bandwidths. Visually, there is large evidence of a discontinuity at zero for each window from both the difference in means and higher order polynomial fits. This graphical evidence lends credence to our claim that beating the spread is a causal treatment to voter beliefs. Since there is strong evidence for balance on both sides of the discontinuity from our balance tables and SUR approach, it is plausible to think that the discontinuity graphed in these figures is the result of a change in the running variable, and not a change in another covariate. Furthermore, these figures provide evidence of a “trigger” for confirmatory bias, which is consistent with our model.

## B Robustness Checks: Inclusion of Controls

Inclusion of control variables should not significantly affect our estimates of confirmatory bias, as treatment assignment should be unrelated to the values of the controls. Indeed, with balanced established the inclusion of additional covariates is likely an over-control which would bias the estimates of confirmatory bias. Although Table 3 and Appendix Tables 1 and 2 provide significant evidence for local balance, we test to see whether our results are robust by including controls that

have been found to affect voter beliefs: winning and losing and beating the spread in previous weeks (see Coleman et al. (2010)). In Appendix Table 5, we estimate our RD design with the inclusion of these controls for all five windows.

We find evidence that, even after including controls for other determinants of voter beliefs, there is still evidence of confirmatory bias. Column 1 of Appendix Table 5 provides our estimates for a basic linear regression of voter beliefs on beating the spread and the controls. In the three, four, and four and one-half point windows, beating the spread is significant at the 90 percent level, with estimates of 15.48 points, 17.25 points, and 20.80 points, respectively. These results are equivalent to between 18 to 20 voters moving a team up one slot in the poll. Columns 2 and 3 include the controls and a fully interacted cubic and quartic polynomial, respectively. Again, we find that beating the spread is a statistically significant determinant of voter beliefs in all five windows, for eight out of ten specifications, and that these results are all significant at least at the 95 percent level. Thus our basic estimates are robust to both inclusion of controls and higher-order specifications designed to account for potential misspecification of the belief function, even though their inclusion is an over-control which is likely to bias estimates of confirmatory bias.

## C Robustness Checks: Constant Bias Across Windows

We also test to see whether or not there is constant bias in our two and one-half, three, four, and four and one-half point windows. To test for constant bias over time, we reproduce our basic estimates in (12) for these windows. Our results are contained in Appendix Table 6. In this table, we find robust evidence across windows and specifications indicating that the amount of bias over time is not changing. We do not find that the interaction term,  $\gamma$ , is distinguishable from zero in any of our windows or specifications. Beating the spread, however, remains significant in nearly every specification. Based on these results, we do not find any evidence that voters react more significantly to an ambiguous signal if it is received later. We find evidence that belief polarization, as described in our model, comes from the aggregation of many altered signals over time, rather than responding more strongly to weak signals at later times.

Appendix Table 1: Balance of Covariates: Within Two and one-half Points

Variable	Below Zero	Above Zero	Difference	SUR Estimate
	(1)	(2)	(3)	(4)
Week of Season	8.622 (2.995)	8.204 (3.119)	-.418 [-1.148]	.603 (3.938)
Score of Team	27.661 (11.105)	30.783 (12.341)	3.122** [1.236]	9.124 (15.295)
Score of Opponent	16.205 (9.294)	17.032 (9.326)	.827 [.744]	-13.775 (11.973)
AP Rank Before Game	10.307 (6.031)	10.955 (6.696)	.648 [.857]	.257 (8.262)
AP Rank After Game	10.551 (6.499)	10.841 (6.781)	.290 [.367]	.051 (8.595)
Wins of Opponent (By Game)	3.756 (2.305)	3.350 (2.312)	-.406 [-1.474]	-4.844 (2.969)
Losses of Opponent (By Game)	2.512 (1.959)	2.650 (2.072)	.138 [.575]	4.739* (2.603)
Win	.827 (.380)	.879 (.327)	.052 [1.220]	.157 (.454)
Home	.551 (.499)	.522 (.501)	-.029 [-.486]	-.218 (.649)
<i>N</i>	127	157	284	284

The first and second columns contain means with standard deviations below. The third column contains t-statistics from a Welch's t-test of a difference in means in the covariates, since the two samples have unequal variances. The fourth column contains the coefficient estimates from a system of seemingly unrelated regressions, where each equation in the system consists of a regression of the variable in the first column on an indicator that assumes a value of one if the team beat the spread and the interaction terms in the running variable. In this column, we test to see if the coefficient on the indicator is zero, and report standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Appendix Table 2: Balance of Covariates: Within Three Points

Variable	Below Zero	Above Zero	Difference	SUR Estimate
	(1)	(2)	(3)	(4)
Week of Season	8.238 (3.162)	8.332 (3.141)	.094 [.291]	.241 (5.249)
Score of Team	28.854 (11.061)	30.898 (12.096)	2.144* [1.723]	1.996 (12.066)
Score of Opponent	16.914 (9.397)	17.133 (9.301)	.219 [.228]	-7.524 (9.686)
AP Rank Before Game	10.265 (6.176)	10.867 (6.783)	.602 [.906]	7.460 (6.775)
AP Rank After Game	10.341 (6.481)	10.577 (6.776)	.236 [.347]	6.231 (6.909)
Wins of Opponent (By Game)	3.616 (2.418)	3.577 (2.445)	-.039 [-.016]	-4.310* (2.501)
Losses of Opponent (By Game)	2.422 (1.955)	2.582 (2.053)	.16 [.779]	1.136 (2.094)
Win	.837 (.370)	.883 (.323)	.046 [1.290]	.128 (.362)
Home	.605 (.499)	.526 (.501)	-.079 [-1.578]	-.291 (.515)
<i>N</i>	185	196	381	381

The above table contains descriptive statistics and estimates of the difference in the variables for teams that have a difference of less than three and a half points between the actual margin of victory and the point spread. The first and second columns contain means with standard deviations below. The third column contains t-statistics from a Welch's t-test of a difference in means in the covariates, since the two samples have unequal variances. The fourth column contains the coefficient estimates from a system of seemingly unrelated regressions, where each equation in the system consists of a regression of the variable in the first column on an indicator that assumes a value of one if the team beat the spread and the interaction terms in the running variable. In this column, we test to see if the coefficient on the indicator is zero, and report standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Appendix Table 3: Balance of Covariates: Within Four Points

Variable	Below Zero	Above Zero	Difference	SUR Estimate
	(1)	(2)	(3)	(4)
Week of Season	8.276 (3.170)	8.405 (3.195)	.129 [.074]	-2.414 (2.210)
Score of Team	28.480 (10.944)	31.103 (11.889)	2.623** [2.084]	-3.610 (7.993)
Score of Opponent	17.146 (9.254)	16.679 (9.422)	-.467 [.168]	2.734 (6.505)
AP Rank Before Game	10.179 (6.315)	10.633 (6.651)	.454 [.610]	6.423 (4.533)
AP Rank After Game	10.256 (6.471)	10.206 (6.595)	-.05 [-.106]	5.852 (4.554)
Wins of Opponent (By Game)	3.699 (2.497)	3.641 (2.578)	-.058 [-.268]	-2.252 (1.767)
Losses of Opponent (By Game)	2.398 (1.861)	2.546 (1.993)	.148 [1.146]	.218 (1.350)
Win	.833 (.373)	.889 (.314)	.046*** [2.354]	.139 (.241)
Home	.618 (.486)	.527 (.500)	-.091** [-2.079]	-.522 (.345)
<i>N</i>	246	262	508	508

The first and second columns contain means with standard deviations below. The third column contains t-statistics from a Welch's t-test of a difference in means in the covariates, since the two samples have unequal variances. The fourth column contains the coefficient estimates from a system of seemingly unrelated regressions, where each equation in the system consists of a regression of the variable in the first column on an indicator that assumes a value of one if the team beat the spread and the interaction terms in the running variable. In this column, we test to see if the coefficient on the indicator is zero, and report standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Appendix Table 4: Balance of Covariates: Within Four and a Half Points

Variable	Below Zero	Above Zero	Difference	SUR Estimate
	(1)	(2)	(3)	(4)
Week of Season	8.327 (3.205)	8.457 (3.152)	.130 [.479]	-2.564 (1.965)
Score of Team	28.543 (11.167)	31.35 (12.023)	2.807*** [2.836]	.519 (7.226)
Score of Opponent	17.502 (9.507)	16.632 (9.363)	-.87 [-1.080]	6.054 (5.845)
AP Rank Before Game	10.126 (6.274)	10.714 (6.638)	.588 [.364]	3.001 (4.034)
AP Rank After Game	10.271 (6.415)	10.236 (6.561)	-.035 [-.063]	2.862 (4.041)
Wins of Opponent (By Game)	3.717 (2.512)	3.643 (2.554)	-.074 [-.342]	-2.353 (1.572)
Losses of Opponent (By Game)	2.431 (1.959)	2.571 (2.024)	.14 [.824]	.262 (1.239)
Win	.818 (.387)	.893 (.31)	.075** [2.500]	.113 (.217)
Home	.613 (.488)	.536 (.499)	-.077* [-1.828]	-.550* (.307)
<i>N</i>	269	280	549	549

The first and second columns contain means with standard deviations below. The third column contains t-statistics from a Welch's t-test of a difference in means in the covariates, since the two samples have unequal variances. The fourth column contains the coefficient estimates from a system of seemingly unrelated regressions, where each equation in the system consists of a regression of the variable in the first column on an indicator that assumes a value of one if the team beat the spread and the interaction terms in the running variable. In this column, we test to see if the coefficient on the indicator is zero, and report standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$



Appendix Table 5: Regression Discontinuity Estimates of Confirmatory Bias: Controls Included

	(1)	(2)	(3)
Panel 1: RD Implementation, Two and One-Half Point Window			
Beating the Spread	9.396 [11.444]	210.925*** [33.956]	-184.798 [142.859]
N	284	284	284
R-squared	.474	.479	.480
Panel 2: RD Implementation, Three Point Window			
Beating the Spread	15.476 [10.149]	199.376*** [45.952]	165.193** [59.761]
N	381	381	381
R-squared	.422	.430	.433
Panel 3: RD Implementation, Three and One-Half Point Window			
Beating the Spread	18.277* [9.496]	140.760*** [38.028]	247.699*** [44.644]
N	417	417	417
R-squared	.397	.405	.407
Controls?	X	X	X
Quadratic in $\theta$ ?		X	X
Cubic in $\theta$ ?		X	X
Quartic in $\theta$ ?			X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. We include controls for whether a team wins or loses and whether or not a team beat the spread in previous weeks. Specifications (2) and (3) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .

Appendix Table 5: Regression Discontinuity Estimates of Confirmatory Bias: Controls Included (continued)

	(1)	(2)	(3)
Panel 4: RD Implementation, Four Point Window			
Beating the Spread	17.247** [7.988]	118.938*** [41.600]	219.004*** [45.032]
N	508	508	508
R-squared	.409	.415	.417
Panel 5: RD Implementation, Four and One-Half Point Window			
Beating the Spread	20.801** [7.988]	54.026 [44.678]	246.398*** [55.184]
N	549	549	549
R-squared	.420	.423	.427
Controls?	X	X	X
Quadratic in $\theta$ ?		X	X
Cubic in $\theta$ ?		X	X
Quartic in $\theta$ ?			X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. We include controls for whether a team wins or loses and whether or not a team beat the spread in previous weeks. Specifications (2) and (3) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .

Appendix Table 6: Testing for Constant Bias over Time

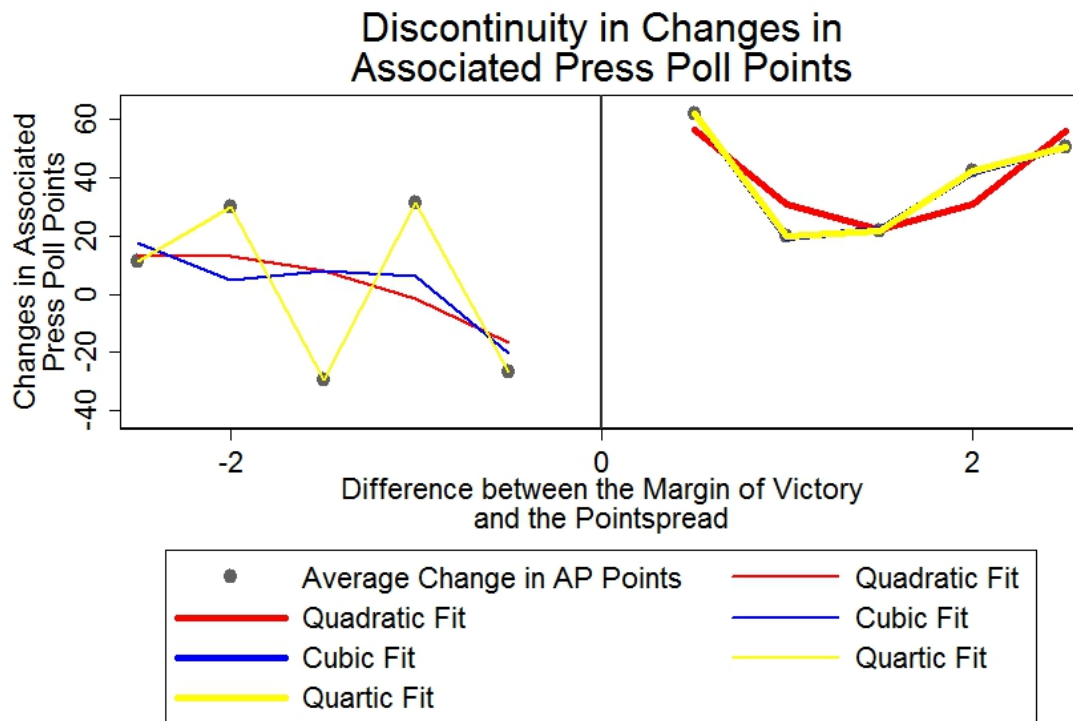
	(1)	(2)	(3)	(4)
Panel 1: Two and One-Half Point Window				
Beating the Spread	32.086 [20.281]	148.579*** [42.495]	280.073 *** [67.891]	1046.65 *** [12.009]
Beating the Spread*Week of Season	-.921 [2.062]	-.508 [2.184]	-.437 [2.191]	-.438 [2.199]
N	284	284	284	284
R-squared	.006	.013	.015	.023
Panel 2: Three Point Window				
Beating the Spread	37.972* [18.583]	118.237*** [37.882]	241.646 *** [67.891]	445.646 *** [126.533]
Beating the Spread*Week of Season	-1.066 [1.662]	-1.288 [1.819]	-1.049 [1.869]	-1.076 [1.883]
N	381	381	381	381
R-squared	.005	.017	.019	.020
Panel 3: Four Point Window				
Beating the Spread	34.361** [14.753]	63.819 [40.649]	160.181 *** [45.617]	307.817 *** [67.042]
Beating the Spread*Week of Season	-.049 [1.293]	-.263 [1.327]	.116 [1.457]	.115 [1.440]
N	508	508	508	508
R-squared	.012	.016	.022	.023
Quadratic in $\theta$ ?		X	X	X
Cubic in $\theta$ ?			X	X
Quartic in $\theta$ ?				X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. Specifications (2)-(4) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .

Appendix Table 6: Testing for Constant Bias over Time (continued)

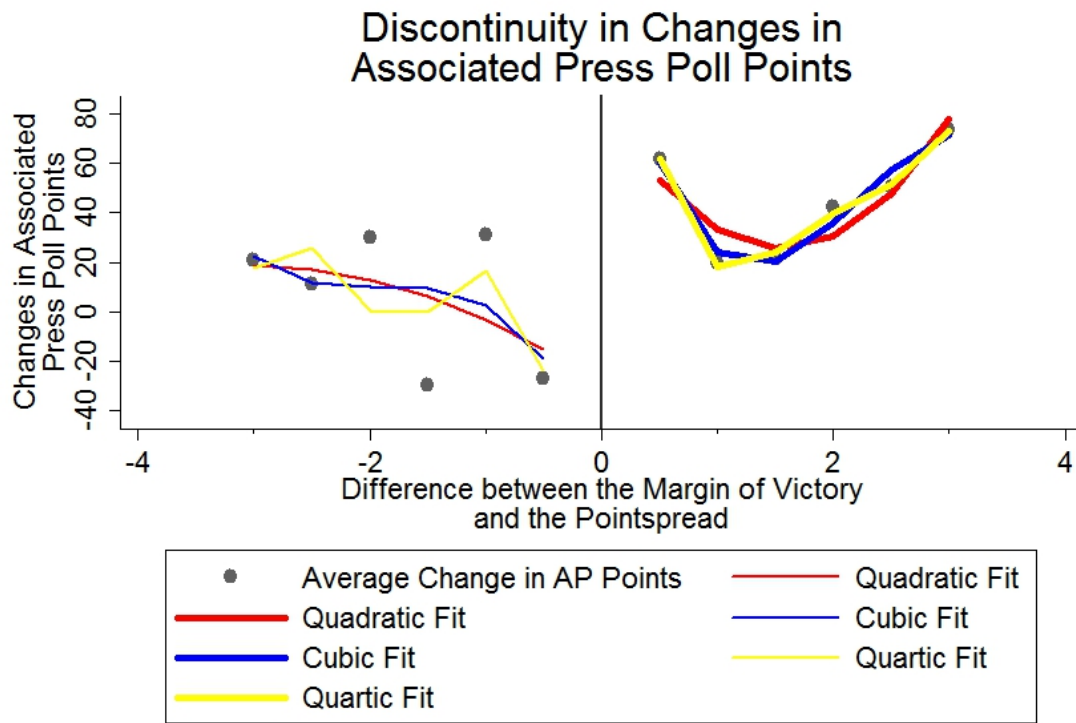
	(1)	(2)	(3)	(4)
Panel 4: Four and One-Half Point Window				
Beating the Spread	41.867** [15.476]	83.341** [35.859]	94.062 * [52.472]	313.402 *** [66.926]
Beating the Spread*Week of Season	.079 [1.238]	-.355 [1.293]	-.223 [1.337]	-.068 [1.354]
N	549	549	549	549
R-squared	.019	.030	.031	.037
Quadratic in $\theta$ ?		X	X	X
Cubic in $\theta$ ?			X	X
Quartic in $\theta$ ?				X

The dependent variable is the change in AP Poll Points from the start to the end of week.  $\theta$  is the running variable, which is the difference between the margin of victory and the betting line. Specifications (2)-(4) include a fully interacted polynomial in  $\theta$  in order to allow the conditional mean function to vary on both sides of the discontinuity. We cluster our standard errors on  $\theta$  to account for potential misspecification of the conditional mean function. Standard errors are in brackets, with significance levels: \*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$ .



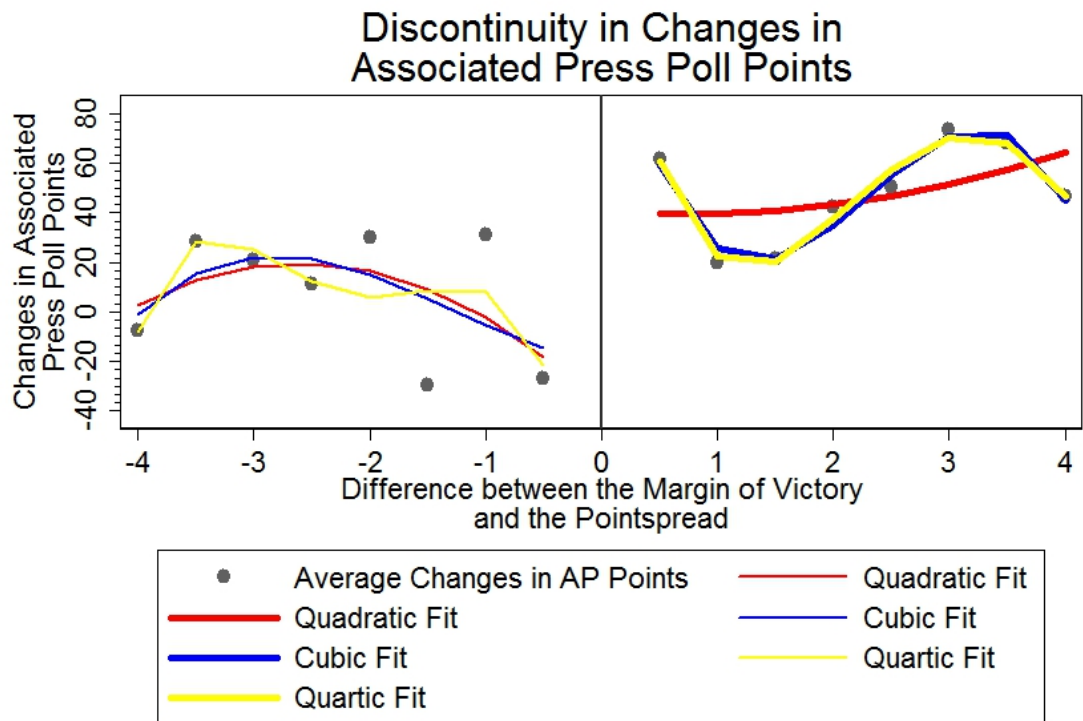
Points are fitted values.

Appendix Figure 1: Belief Dynamics at the Discontinuity, Higher Order Polynomial Fits, 2.5 Point Window

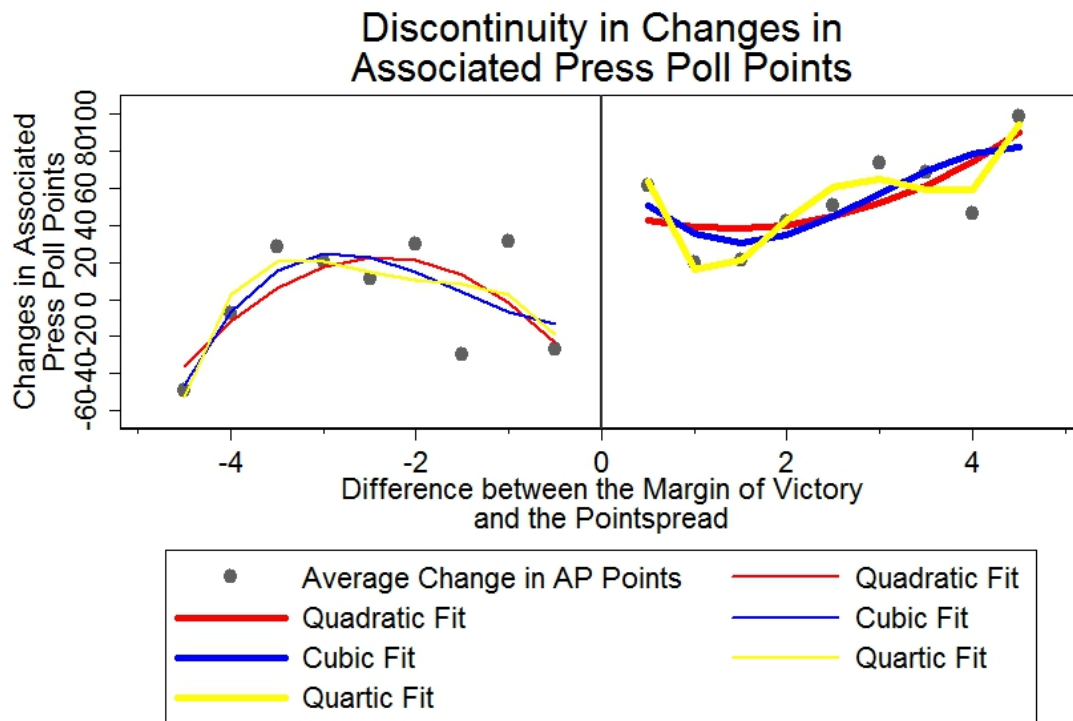


Points are fitted values.

Appendix Figure 2: Belief Dynamics at the Discontinuity, Higher Order Polynomial Fits, Three Point Window



Appendix Figure 3: Belief Dynamics at the Discontinuity, Higher Order Polynomial Fits, Four Point Window



Points are fitted values.

Appendix Figure 4: Belief Dynamics at the Discontinuity, Higher Order Polynomial Fits, 4.5 Point Window