2014

Discrimination and Information: Geographic Bias in College Basketball Polls

Andrew Nutting
Bryn Mawr College, anutting@brynmawr.edu

DISCRIMINATION AND INFORMATION: GEOGRAPHIC BIAS IN COLLEGE BASKETBALL POLLS

Andrew W. Nutting
Department of Economics
Hamilton College
Clinton NY 13323
Phone: (315) 859-4074
March 2014

ABSTRACT— Voters in the Associated Press college basketball poll vote own-state teams and teams that are fewer miles away to higher rankings than other teams, especially at the bottom of their ballots. Game outcome data show evidence that teams that are fewer miles away are underrated—not overrated—by pollsters, especially at the top of their rankings, perhaps because pollsters fear accusations of geographic bias. When controlling for distance between pollsters and teams, there is some evidence that pollsters overrate local-conference teams at the top of their ballots, but more properly rate them the bottom of their ballots.

I. Introduction

According to economic theory, discrimination is either taste-based [Becker 1957] or information-based, i.e. statistical [e.g. Aigner and Cain 1977]. Taste-based discrimination occurs when economic agents possess a dislike of certain groups and potentially forgo economic benefits such as profits or wages to avoid them. Information-based discrimination involves economic agents possessing imperfect information about individuals, and therefore assuming they possess the mean characteristics of their group. There may be other forms of discrimination as well. Wolfers [2006] uses the term "mistake-based" discrimination to describe a group being treated differently because its mean characteristics are not correctly known.[1]

This paper examines discrimination in the context of college basketball polls with the goal of distinguishing whether discrimination is taste-based or information-based. Using one year's worth of college basketball polls and games, it first determines whether, in a given week, pollsters rank teams differently based on their geographic proximities to teams. Coleman et al [2010] used this technique to show that media voters in the 2007 Associated Press College Football Poll voted teams from their home state and teams from conferences that include own-state teams to higher rankings than other teams.[2] Fixed effects estimations show that pollsters vote own-state teams and teams that are fewer miles away to better ranks than other teams, especially at the bottom of their polls.

Second, this paper examines whether pollsters rank teams in close geographic proximity more accurately or less accurately than other teams. Assuming that *ceteris paribus* pollsters should aim to create a poll where better-ranked teams defeat worse-ranked teams, a finding that teams were more likely, for example, to lose when pollsters voted them to higher ranks would indicate that pollsters engage in taste-based discrimination favoring local teams. If instead empirical estimations showed that teams were more likely, for example, to win when own-state pollsters placed them to better ranks, it would suggest that pollsters possessed better information about teams in closer geographic proximity and therefore *more accurately ranked* them. In such a case, own-state and more-nearby teams voted by pollsters to higher ranks would win more often than other teams of the same rank. Ross, Larson, and Wall [2012], using a similar methodology but not utilizing data on individual pollsters, find that college football pollsters overrate teams from the Mid-American Conference and underrate teams from the Southeastern Conference and former Pac-10 Conference. Estimations show that pollsters rank own-state teams and teams that are fewer miles away to higher rankings than other teams. The effect is especially strong at the bottom of the polls. Estimations show no evidence that pollsters overrate teams in close geographic proximity. In fact, evidence suggests pollsters more accurately rank own-state teams and may actually underrate teams that are fewer miles away, especially at the top of their polls. There is also some evidence that pollsters overrate teams in local conferences at the top of their polls, but correctly award them better ranks at the bottom of their polls.

This paper adds to the growing literature examining the sources of discrimination. Altonji and Pierret [2001] find that statistical discrimination in labor markets declines over workers' careers because employers become better informed about individual workers' productivity levels. List [2004] shows that discrimination against minority groups (women, racial minorities, and the older-aged) in negotiations over baseball card prices constitutes statistical, not taste-based, discrimination. Levitt [2004] finds that contestants on the television game show *The Weakest Link* exhibit taste-based discrimination against older contestants and statistical discrimination against Hispanic contestants (the latter by believing they are worse players than white contestants). Holzer, Raphael, and Stoll [2006] find that employment discrimination against black males is reduced when employers can conduct criminal background checks, suggesting such discrimination is statistical. Wolfers and Kumar [2008] discern that male analysts significantly under-predict the performance of firms with female CEOs, suggesting taste-based discrimination. Parsons, Sulaeman, and Hamermesh [2011] find that Major League Baseball umpires call more strikes when the pitcher is of the same race, and believe such discrimination is taste-based because it decreases when umpire monitoring, and therefore the costs of discrimination, rise. Gneezy, List, and Price [2012] show that people are more likely to engage in taste-based discrimination against those who have characteristics—such as obesity and homosexuality—that are believed to be controllable.

The remainder of this paper is organized as follows. Section II discusses the dataset and details why this paper examines college basketball instead of college football. Section III empirically determines whether geography affects basketball pollsters' rankings. Section IV determines whether there is a relationship between pollster/team proximity and pollsters' accuracy in ranking. Section V concludes.

II. College Basketball Polls

The Associated Press (AP) has conducted a poll of the nation's top college basketball teams since 1949, according to data at the website *collegepollarchive.com*. Since the 1989-90 season it has publicly ranked the Top 25 teams. Rankings are determined by "Borda count." Each week pollsters rank their best 25 teams in the country. The top team on each pollster's ballot receives 25 points, the second-place team 24, and so on through the #25 team, which receives 1 point. Points are summed across pollsters, and the team with the most total points is ranked #1, the team with the second-most points is ranked #2, and so on, to #25. Teams that receive points but are not in the Top 25 are named along with their point totals, but not officially ranked.

In the 2009-10 season, the AP poll was released at 19 different points in the season (once weekly, plus a preseason poll], and each poll consisted of 65 pollsters. All pollsters are listed in Appendix Table 1,[3] and each pollster's complete poll for each week is available at the website *pollspeak.com*. Pollsters represented 40 different states plus the District of Columbia. Division I men's college basketball in 2009-10 consisted of 347 teams, each of which could be ranked in any pollster's ballot in any given week. Date and final score data for all men's college basketball games for 2009-2010 were taken from the website *espn.com*.

There are interesting differences between analyzing college basketball polls and college football polls. Compared to college football polls, there may be less potential benefit to pollsters of engaging in unabashed rent-seeking vis-à-vis favoring local teams. In college football, where the Coaches Poll has recently helped determine who plays in the national championship and other high-ranking bowl games, the influence Coaches Poll voters have over the football season's outcomes provides strong incentives to vote preferred teams to higher rankings [Mandel 2005; Mirabile and Witte 2010].[4] Though the AP Football Poll does not directly determine BCS standings, it is possible that voters in the AP and Coaches Polls affect each other [Stone and Zafar, forthcoming], and AP Poll rankings could be affected by Coaches Poll voters' strong incentives to vote certain teams to higher ranks. In college basketball, incentives towards ballot manipulation are smaller, because the single-elimination postseason tournament means that weekly AP College Basketball polls do not affect which teams compete for the national championship.[5] Thus college basketball polls may be a better indicator of pollsters' *implicit*—rather than unabashedly rent-seeking—geographic biases than football polls.[6] Implicit discrimination, defined by Bertrand, Chugh, and Mullainathan [2005] as being "unintentional and outside the discriminator's awareness," is a growing field of empirical economic research [Price and Wolfers 2010; Fowder, Kadiyall, and Prince 2012]. Bertrand, Chugh, and Mullainathan [2005] show that implicit biases can have large repercussions.

Another reason for using college basketball polls are that college basketball seasons contain more weeks and games than college football seasons, providing more observations for empirical study.[7]

III. Evidence of Pollster Geographic Bias

a. Empirical Strategy

Where i is individual basketball team, p is individual pollster, and t is week of the season, pollster geographic bias is determined through estimations of the equation

$$B_{ipt} = \phi X_{ip} + \kappa_{it} + \sigma_p + \varepsilon_{ipt}. \tag{1}$$

B_{ipt} is equal to the number of Borda count points awarded to team i in week t by pollster p. Since B_{ipt} is upper-bounded at 25 and lower-bounded at 0, estimations of Equation (1) are performed via a two-sided maximum likelihood Tobit estimation censored at both an upper limit of 25 and a lower limit of 0. The Tobit estimation requires that the error term ε_{ipt} be normally distributed.

X_{ip} is a vector of controls measuring the geographic relationship between team i and pollster p. It consists of three variables, all used in Coleman et al [2010]: *STATE_MATCH*, a dummy variable equal to 1 if i and p are in the same state;[8] *CONFERENCE_MATCH*, a dummy variable equal to 1 if i is in a different state than p, but is in the same conference as a team located in p's state; and *DIST*, a variable measuring the miles in thousands between team i's campus and pollster p's employer.[9] Throughout this paper, teams where *CONFERENCE_MATCH* is 1 are referred to as "conference-affiliated" teams.

The control κ_{it} in Equation (1) is a fixed-effects control representing team it, i.e. each team separately for each week. κ_{it} accounts for the average number of Borda count points team i receives from pollsters in week t. It renders superfluous any controls that are constant within team it, such as win-loss record [Coleman et al 2010], national television exposure [Campbell, Rogers, and Finney 2007], recent performance [Logan 2007], quality of opposition, presence of injuries, historical reputation of program, etc. The inclusion of κ_{it} in Equation (1) also accounts for potential confounding relationship between pollsters, teams, and geography. For example, since the Associated Press aims for geographic diversity in its pollsters, it may oversample regions of the country (such as the Rocky Mountain area) that have relatively few high-quality basketball teams. This would create a spurious negative relationship between pollster geographic proximity and Borda count points unless κ_{it} is included.

Since estimations of Equation (1) are Tobits, teams ranked by no pollsters in week *t* have unidentified team/week effects in κ_{it} and contribute no information to estimations of the other parameters in Equation (1). Therefore estimations are restricted to observations of teams ranked by at least one pollster in week *t*. If team *it* is ranked by some pollsters, but not all of them, *it* is represented in the dataset with 65 observations (one for each pollster), with B_{ipt} values equal to 0 for pollsters who omit them from their Top 25.

σ_p in Equation (1) is a fixed-effects vector representing each pollster. Since each pollster awards the same number of Borda count points in each week—25 to the top team, 24 to the #2 team, etc—σ_p only accounts for *p*'s probability of awarding zero points to team *it*. That is, σ_p controls for a pollster's probability to omit teams that are ranked by other pollsters. σ_p does not vary by week, so a pollster's propensity of omitting marginal teams from his[10] ballot is assumed to remain constant over the entire season.[11] Since each pollster's Borda count points all sum to the same amount in the same week, standard errors are correlated within pollster [Hausman and Rudd 1987; Stone and Zafar, forthcoming]. Thus Equation (1) standard errors are clustered by pollster-week.

b. Summary Statistics and Results

Table 1 Panel A shows summary statistics of B_{ipt}, the three X_{ip} variables, and *RANKED*, a dummy variable equal to 1 if team *i* is ranked in poll *pt*. Since there are 19 polls in the season, 25 teams ranked per pollster, and 65 pollsters per poll, there are (19 * 25 * 65) = 30,875 observations with positive B_{ipt} values. Panel A contains 59,150 observations because it includes observations where team *it* is unranked by pollster *p* but ranked by at least one other pollster. In Column 1, the mean number of Borda count points is fewer than 7 and team *it* is unranked in almost half of observations. The mean value for *STATE_MATCH* is 0.031, the mean value for *CONFERENCE_MATCH* is 0.206, and the average team is located almost 900 miles from the average pollster. When omitting unranked observations (Panel B), the mean values of *STATE_MATCH* and *CONFERENCE_MATCH* increase slightly and the average team becomes about 30 miles closer to the average pollster. This suggests that pollsters are more likely to place closer teams to higher ranks.

Table 2 shows results from estimations of Equation (1). Coefficients can be interpreted as marginal effects. Column 1 shows results when omitting the control for *DIST*.[12] Pollsters award 1.06 additional statistically significant points to teams from their home state, and 0.3 additional statistically significant points to conference-affiliated teams. When adding *DIST* as a control (Column 2), *DIST* is significantly negative, indicating that a pollster awards a team 2400 miles away 1 fewer point, *ceteris paribus*, than a nearby team.[13] *STATE_MATCH* remains significant but falls from 1.06 to 0.71, indicating that approximately one-third of the one-point bonus that pollsters give own-state teams is because of their close proximity. *CONFERENCE_MATCH* becomes insignificant when controlling for distance, indicating that pollsters only vote conference-affiliated teams to higher rankings because they are closer in distance than other teams.

Columns 3-5 repeat Column 2 on subsamples of the data. Column 3 shows coefficients from a fixed-effects logit estimation where the dependent variable is 1 if team *i* is ranked in poll *pt* and 0 if not. Teams ranked by every pollster are dropped from the sample, because they are perfectly classified in the logit fixed-effects estimation. All three coefficients are significant, showing that pollsters are significantly more likely to place own-state teams, conference-affiliated teams, and closer teams in their Top 25 than other teams.

Column 4 shows how geography affects the bottom half of pollsters' ballots. The dependent variable, *POINTS*, is censored at an upper limit of 13 Borda Count points (equivalent to a rank of 13) and a lower limit of 0 (unranked). Teams not ranked between 13th and 25th in any week *t* ballot are removed from the estimations. *CONFERENCE_MATCH* is insignificant in Column 4, but *STATE_MATCH* and *DIST* are significantly positive. In the bottom half of their ballots, pollsters award own-state teams 1.2 more Borda count points and award 1 fewer point for every 1500 additional miles away a team is.

Column 5 shows how geography affects the top half of pollsters' ballots. In this estimation, the dependent variable is censored at an upper limit of 25 (equivalent to a rank of 1) and a lower limit of 12 Borda Count points (equivalent to a rank of 14) instead of 0. Teams that fail to receive 13 or more points in at least one week *t* ballot are dropped from the estimation. In the top half their ballots, pollsters award a team 2900 miles away approximately 0.5 fewer points than a team 0 miles away. Interestingly, pollsters award conference-affiliated teams 0.12

significantly fewer Borda count points. There is no significant difference in ranking for own-state teams.

Columns 3-5 indicate that preferences for teams in closer geographic proximity are stronger at the bottom of college basketball polls than at the top. That the preference for proximity is more intense at the bottom of ballots may be because pollsters' discernment of team quality is significantly worse at lower ranks than at higher ranks [Nutting 2011; Stone 2013], and pollsters may allow their geographic biases to play larger roles when ordinal rankings are less clear. Another factor is that negative repercussions of bias may be lessened at lower ranks. Pollster John Feinstein, for example, has openly admitted that he reserves the #25 spot for low-profile local teams when compiling his ballot [Steinberg 2007].

The use of Borda count as a dependent variable may be problematic, because its cardinal and spatially autoregressive nature—each pollster is constrained to offering only 325 points total in a given week—results in both direct effects (i.e., ranking team *i* higher) and indirect effects (i.e., ranking other teams lower because team *i* is ranked higher) being summed into one coefficient in a Tobit estimation [LeSage and Pace 2009]. This may render the marginal effects in Table 2 biased away from zero, showing more of a relationship between geography and rankings than there actually is. To placate fears that the significance of the coefficients in Table 2 reflect only this endogeneity, I re-estimate Equation (1) using an ordered probit maximum likelihood approach. Ordered probit estimations are less subject to spatial autoregressive bias than Tobits because they treat rankings within a given week as ordinal and not cardinal. Results in Appendix Table 2 show that marginal effects in the ordered probit estimations of Equation (1) are smaller than in the Tobit estimations, as anticipated. Importantly, signs and significances are unchanged, suggesting the significant relationship between geography and ranking uncovered in Table 2 is not solely due to spatial autoregressive bias.

To further investigate the relationship between geography and pollster behavior, Figures 1-4 show nonparametric effects of geography on Borda count points using a Tobit analysis. For each team where *STATE_MATCH*=1, Equation (1) is used to create a linear predicted value of *B*. The positive own-state effect is subtracted, so that

$$\hat{B}'_{ipt} = \hat{\phi} X_{ip} + \hat{\kappa}_{it} + \hat{\sigma}_{p} - \hat{\phi}_{s} STATE\_MATCH_{ip} . \quad (2)$$

This value is used to create an expected value $E(B'_{ipt})$ between 0 and 25 [Wooldridge 2002].[14] Then, for own-state teams only, and where B_{ipt} is the actual number of Borda count points, the state effect $\hat{v}_{ipt}$ is created, where

$$\hat{v}_{ipt} = B_{ipt} - E(B'_{ipt}) . \qquad (3)$$

Figure 1 shows probability density functions of $\hat{v}_{ipt}$ for own-state teams in both the full-sample estimation and the bottom-half-sample estimation.[15] Values of $\hat{v}_{ipt}$ are grouped into units of 0.5 and the x-axis values show group minimum values, e.g. "0" refers to having $\hat{v}_{ipt}$ between 0 and +0.5. The plurality of own-state observations have a $\hat{v}_{ipt}$ between -0.5 and 0. Over 85 percent of the time in the full-sample estimation and over 94 percent of the time in the bottom-half estimation, a $\hat{v}_{ipt}$ in this range indicates a team ranked by someone into the Top 25, but not the own-state pollster in this particular observation.

That said, the Figure 1 pdfs are skewed to the right, especially in the tails. In the full sample, pollsters are over three times more likely to give own-state teams 2.5-3 more Borda count points than 2.5-3 fewer points. They are over four times more likely to give them 4-4.5 more points than 4-4.5 fewer points. The distribution in the bottom-half sample is similarly skewed.

Importantly, Figure 1 shows that though pollsters vote own-state team to significantly higher ranks (Table 2), they also frequently vote own-state teams to lower ranks than other pollsters. Pollsters, therefore, do not blindly give extra Borda count points to teams in close geographic proximity, but rather do so selectively.

Figure 2 shows nonparametric distance effects in the full-sample estimation. (The distance effect $\hat{\delta}_{ipt}$ is constructed analogously to the state effect above.) Figure 2 also shows that, even though pollsters vote more-nearby teams to significantly higher ranks, they can and do vote them to lower ranks as well. In Figure 2, the three pdfs respectively show the full-sample distributions of $\hat{\delta}_{ipt}$ for teams that are 0-100 miles from the pollster, 600-700 miles from the pollster, and 1500-1600 miles from the pollster. (These distances are somewhat arbitrary, and respectively reflect nearby teams, average-distance-away teams, and far-away teams.) Figure 2 shows that $\hat{\delta}_{ipt}$ is more likely to be between 0 and +1 for teams 0-100 miles away than teams that are farther away. Teams 1500-1600 miles away are more likely to see $\hat{\delta}_{ipt}$ be between -1.5 and -1 than closer teams. The tails in Figure 2 are fairly similar for all three groups.

Figure 3 and Figure 4 repeat Figure 2 for, respectively the bottom-half and top-half estimations. Figure 3 shows that teams 0-100 miles away are much more likely to get 0-0.5 extra Borda Count points than teams 600-700 or 1500-1600 miles away, and teams 1500-1600 miles away are noticeably more likely to end up with $\hat{\delta}_{ipt}$ between -2 and -1.5. Figure 4 shows some evidence that teams 0-100 miles away are more likely to see $\hat{\delta}_{ipt}$ between +0.5 and +1.5 than other teams, and teams 1500-1600 miles away are less likely to see $\hat{\delta}_{ipt}$ between +1 and +2.5 than other teams.

IV. Pollster accuracy and the sources of geographic discrimination

a. Estimation strategy

The next estimations examine whether pollsters' accuracy in ranking teams varies with their geographic proximity to teams. The assumption of these estimations is: *pollster* p *accurately ranks two teams if, when the two teams play each other in week* t*, the team with the better rank in poll* pt *defeats the team with the worse rank in poll* pt (poll *pt* being the poll in effect the day the game is played). This is what Coleman [2005] calls "predictive accuracy."[16] Where *p* is pollster, *t* is week, *g* is individual game, and *b* and *w* are the better-ranked and worse-ranked teams in game *g*, logit estimations take the form

$$WIN_{gpt} = \varphi_p + \eta_{gt} + \gamma_b X_{bgpt} + \gamma_w X_{wgpt} + \tau_b RANK_{bgpt} + \tau_w RANK_{wgpt} + e_{gpt}. \qquad (4)$$

WIN_{gpt} is a dummy variable equal to 1 if, in game *g* and poll *pt*, the better-ranked team defeats the worse-ranked team. WIN_{gpt} is identified if the better-ranked team is ranked in the Top 25 of poll *pt*. (The worse-ranked team can be unranked.) φ_p is a vector of fixed effects controlling for each pollster's season-long predictive accuracy. η_{gt} is a vector of fixed effects controlling for game *g* in week *t*. η_{gt} controls for the average pollster's accuracy in predicting game *g*, and controls for factors that affect game outcomes, such as which team is playing at home, whether certain players are injured, etc. X_{bgpt} and X_{wgpt} are vectors of variables capturing the geographic relationship between pollster *p* and, respectively, the better-ranked team and the worse-ranked team in game *g*. Their coefficients show whether pollsters' accuracy changes when ranking own-state, conference-affiliated, and more-nearby teams. Separate controls for X_{bgpt} and X_{wgpt}

permit asymmetries in the effects of pollsters' geographic proximity to the better-ranked team and the worse-ranked team.[17]

$RANK_{bgpt}$ and $RANK_{wgpt}$ are, respectively, controls for the actual ordinal rankings of the better-ranked and worse-ranked team in poll *pt*. If the better-ranked team is ranked #1, for example, it presumably has a better chance of winning game *g* than were it ranked #11. Similarly, if the worse-ranked team is ranked #4, it presumably has a better chance of winning than were it ranked #16. The *RANK* controls in Equation (4) are both quadratic in form since poll accuracy is worse at lower ranks [Nutting 2011; Stone 2013]. When the worse-ranked team is unranked, $RANK_{wgpt}$ is represented by a dummy variable. Since geographic biases are manifested via changes in ordinal rank, the *RANK* controls in Equation (4) may be endogenous. Therefore estimations of Equation (4) are performed both including and excluding them.

Equation (4) estimations are logits, so e_{gpt} takes a logistic distribution. The game fixed effects η_{gt} present perfect classification problems, because if game *gt* is correctly (or incorrectly) predicted by every pollster, its coefficient in η_{gt} is not identified. Thus games that every pollster correctly or incorrectly predicts are dropped from the sample, leaving only games in which there is disagreement among pollsters regarding which of the two teams in game *gt* is better-ranked. Standard errors are clustered by pollster.

Table 3 shows summary statistics for games involving teams ranked in poll *pt*. *WIN* is equal to 1 if the better-ranked team in poll *pt* wins and 0 if it loses. *STATE_MATCH_BETTER* and *STATE_MATCH_WORSE* show whether the pollster resides, respectively, in the better-ranked team's state and the worst-ranked team's state. *CONFERENCE_MATCH_BETTER* and *CONFERENCE_MATCH_WORSE* show respectively whether the better-ranked team and the worst-ranked team are "conference-affiliated" to the pollster. *DIST_BETTER* and *DIST_WORSE* show the distance in thousands of miles between the pollster's place of work and, respectively, the better-ranked and worse-ranked team. *BETTER_RANK* and *WORSE_RANK* show the actual ordinal rankings of the two game *g* teams in poll *pt*, and *WORSE_UNRANKED* is a dummy equal to 1 if the worse-ranked team is unranked.

Panel A shows summary statistics for games where the worse-ranked team can be unranked. The two separate sections of Panel A show, respectively, summary statistics for games included in estimations of Equation (4), and games omitted from Equation (4) because every poll either correctly or incorrectly predicts its outcome. There are 120 different games and

5484 total observations included in the estimation sample, so the average game is observed in almost 46 different pollsters' ballots. When including unranked teams in the sample (Panel A), the better-ranked team wins 53.5 percent of observations. The better-ranked team is in the pollster's own state in 3.7 percent of observations, and the worse-ranked team is in the pollster's own state in 3.0 percent of observations. Approximately 21 percent of both better-ranked and worse-ranked teams are conference-affiliated. The pollster resides an average of 838 miles from the better-ranked team and 875 miles from the worse-ranked team. The better-ranked team's mean rank is 14.8, and the worse-ranked team is unranked in almost half of observations. Panel A shows that over 10 times as many games are omitted from the sample as included in the sample. Omitted games are much more likely to feature an unranked team, and a win for the better-ranked team. The state-match, conference-match, and distance values are not especially different between included and omitted games.

Panel B restricts the sample to games in which both teams are ranked in the Top 25 of poll *pt*. Only 76 games, with 37.5 pollsters per game, are included in the sample, and this comprises 35 percent of all games featuring two ranked teams.[18] Among games included in the sample, the better-ranked team has a ranking of better than 11th and wins almost 60 percent of observations. The *DIST* variables fall slightly, and the *STATE* and *CONFERENCE* variables increase slightly, compared to their Panel A analogs. This tendency, where teams in games where both teams are ranked are in closer proximity to pollsters' residences, again indicates pollsters' geographic bias. Games omitted from the sample feature better-ranked teams that are much more highly ranked, and are much more likely to feature the better-ranked team winning. The geography variables are, once again, fairly similar to those of the estimation sample.

b. Coefficients, discrimination, and information

A critical assumption in this section is that pollsters may possess more information about teams in closer proximity, but also possess information about farther-away teams. In other words, pollsters possess an ability, based on previous experience, to discern what qualities make a good team and a bad team. They are able to watch a local team and determine whether it is, for example, a "Top 5 team" or a "Top 15 team," and knowing this they are able to determine whether the rest of the country has the team overrated or underrated.

Table 4 shows how coefficients from estimations of Equation (4) can indicate the form of discrimination being observed. Column 1 shows signs that should appear if pollsters have unambiguously better information about local teams. If pollsters possess better information and vote own-state teams more accurately than other teams, they would be correct when they have own-state teams ranked higher, and the coefficient on *STATE_MATCH_BETTER* would be positive. They would also be correct when they have the own-state team ranked lower, so the coefficient on *STATE_MATCH_WORSE* would be positive too. The coefficients on the *CONFERENCE* coefficients would be interpreted analogously and the coefficients on the *DIST* variables would be reversed, since *DIST* is inversely related to proximity. If pollsters actually had worse information about teams in close proximity (Column 2), all the signs from Column 1 would be reversed.

If taste-based discrimination caused pollsters to vote own-state teams to higher ranks, own-state teams would lose more often to worse-ranked teams, and the coefficient on *STATE_MATCH_BETTER* would necessarily be significantly negative (Column 3). *STATE_MATCH_WORSE* could be either insignificant or significantly positive. A significantly positive *STATE_MATCH_WORSE* would, though, (in conjunction with a positive *STATE_MATCH_BETTER*) provide strong evidence of taste-based discrimination, because pollsters would be less accurate when voting own-state teams to better ranks and more accurate when voting them to worse ranks. This would indicate that voters sacrifice accuracy in order to discriminate. In the event that pollsters actually underrate own-state teams (Column 4), the coefficient on *STATE_MATCH_WORSE* would be significantly negative, indicating that pollsters are incorrect when voting own-state teams to worse ranks. For both Columns 3 and 4, the *CONFERENCE* coefficients are once again analogous to the *STATE* coefficients, and the *DIST* coefficients are reversed.

c. Baseline results

Table 5 shows results from estimations of Equation (4) when omitting controls for miles of distance between pollsters and teams. (Recall that Table 2 Column 1 showed that own-state and conference-affiliated teams had significantly higher ranks when omitting distance controls.) Column 1 includes games involving unranked teams and omits controls for the ordinal ranks of the two teams. *STATE_MATCH_BETTER*, *CONFERENCE_MATCH_BETTER*, *STATE_MATCH_WORSE*, and *CONFERENCE_MATCH_WORSE* are all positive but insignificant. Results are barely changed when adding *RANK* controls (Column 2).[19] When limiting the sample to games in which both teams are ranked in the Top 25 of poll *pt* (Columns

3-4) all coefficients are again insignificant. Thus there is no evidence in Table 5 that pollsters are more or less accurate when ranking own-state or conference-affiliated teams to higher ranks.

Table 6 repeats Table 5 but separates games into “Conference Games” (between teams in the same conference, Columns 1-4) and “Non-Conference Games” (between teams in different conferences, Columns 5-8). *CONFERENCE_OPPONENT_MATCH* is not identified in Columns 1-4 because its value is identical to that of *CONFERENCE_MATCH*. Columns 1-2 show that, with respect to conference games, pollsters are more likely to be correct when voting an own-state team to a higher ranking. When limiting the sample to games where both teams are ranked, the coefficient on *STATE_MATCH* remains large, but becomes insignificant, presumably due to smaller sample size. Columns 6 and 8 show that, with respect to non-conference games, pollsters are more likely to be correct when voting conference-affiliated teams to higher ranks. Table 6 thus suggests that pollsters do not overrate own-state and conference-affiliated teams, but selectively vote them to higher rankings when their information indicates such teams are better than their national reputations suggest.

Table 7 again uses the full sample as its estimation population, and adds controls for *DIST_BETTER* and *DIST_WORSE* to the estimations in Table 5. In Column 1, *STATE_MATCH_BETTER* and *CONFERENCE_MATCH_BETTER* are again insignificant, so there is no evidence that pollsters overrate own-state and conference-affiliated teams. *CONFERENCE_MATCH_WORSE* is significantly positive, showing that pollsters are significantly more accurate when they rank conference-affiliated teams to lower rankings. This indicates that pollsters place conference-affiliated teams at lower rankings when they know they are of worse quality. *STATE_MATCH_WORSE* is similarly positive, but only significant at the 12% level.

DIST_BETTER is significantly negative, showing that when pollsters rank a more-nearby team to a better rank, the more-nearby team is more likely to win. This suggests that pollsters rank more-nearby teams to significantly higher ranks when they know such teams are of high quality. But the significantly positive coefficient on *DIST_WORSE* indicates that when pollsters rank a more-nearby team to a worse ranking, they are *incorrect* significantly more often. These interesting results show that even though pollsters award more-nearby teams higher rankings than further-away teams (Table 2), they may not rank them as highly as they should, and too frequently award them lower ranks. This may occur because pollsters are concerned that, even if they know a more-nearby team to be of high quality, publicly voting them to a high rank may reek of taste-based bias [Morris 2001]. Also, if local pollsters possess more information about more-nearby teams, they may overreact to such information [Stone 2013] and, for example, punish a local team too severely if it loses or plays poorly in a recent game.

Column 1 results are unchanged when controlling for the ordinal rankings of the two teams (Column 2). Columns 3-4 repeat Columns 1-2 but omit games where the worse-ranked team is unranked in poll *pt*. Results strengthen, and *STATE_MATCH_WORSE* becomes significantly positive, showing pollsters are significantly more accurate when voting own-state teams to lower rankings.

Table 8 repeats the estimations shown in Table 7 separately for conference games and non-conference games. The coefficients on the distance variables are roughly analogous in sign and significance to their full-sample analogs, though they are much stronger in intensity among conference games. This suggests pollsters are especially

c. Extensions

Table 9 examines whether the relationship between geographic proximity and pollster accuracy changes at different ranks. *STATE_MATCH_BETTER*, *CONFERENCE_MATCH_BETTER*, and *DIST_BETTER* are interacted with the ordinal rank of the better-ranked team, while *STATE_MATCH_WORSE*, *CONFERENCE_MATCH_WORSE*, and *DIST_WORSE* are interacted with the ordinal rank of the worse-ranked team. Controls for the ordinal rankings of the two teams are included in all estimations to separately identify the interaction effects from the *ceteris paribus* effects of pollster accuracy at different ranks. For simplicity, games involving unranked teams are removed from these estimations.

Column 1 omits *DIST* controls for the sample. It shows no significant evidence that accuracy involving own-state or conference-affiliated teams varies at different ranks. Column 2 adds the *DIST* controls and their interaction. *DIST_BETTER*BETTER_RANK* is significantly positive and *DIST_WORSE*WORSE_RANK* is significantly negative. Both of these coefficients indicate a weakening relationship between distance and accuracy as pollsters move down their ballots. The relationship discussed earlier—that pollsters are correct when they place more-nearby teams to higher rankings, and incorrect when they do the opposite—is strong at the top of pollsters' ballots, but weak at the bottom of ballots. This may indicate that, if pollsters are fearful of being perceived as biased towards teams that are fewer miles away [Morris 2001], their fear is most palpable when voting high-ranked and high-profile teams. Thus they underrate more-nearby teams most strongly at the top of the polls.

Column 3 shows results when omitting the *STATE_MATCH* controls and counting own-state teams as conference-affiliated teams. *CONFERENCE_MATCH_BETTER** *BETTER_RANK* is significantly positive at the 10% level,[20] while *CONFERENCE_MATCH_WORSE*WORSE_RANK* is negative, though statistically insignificant. Recall also that in Table 2 pollsters ranked conference-affiliated teams lower at the top of their ballots and higher at the very bottom of their ballots. In tandem, the Table 2 and Table 9 results suggest that pollsters rank conference-affiliated teams (including own-state teams) more accurately at the bottom of their ballots, where these teams are more likely to win when ranked higher. At the top of their ballots, even though they conference-affiliated teams to significantly lower ranks in Table 2—Columns 2-3 show that they may indeed still be overrated, i.e. they lose significantly more often when ranked higher and when *BETTER_RANK* is smaller.

Since some researchers have expressed concerns about interaction terms in logit models [Ai and Norton 2003], Columns 4-6 reproduce Columns 1-3 using a fixed effects linear probability model instead of a fixed effects logit model. Coefficients on interactions terms have the same signs, though significance weakens.

Table 10 adds different interaction controls to the Table 6 baseline estimation. *STATE_ INTERACT* is equal to *STATE_MATCH_BETTER*STATE_MATCH_WORSE*. *CONFERENCE_INTERACT* and *DIST_INTERACT* are defined analogously. The bottom of each column shows the sum of the three *STATE* coefficients and the sum of the three *CONFERENCE* coefficients, and the respective p-values of chi-square tests of significance. These show whether pollsters more correctly order own-state or conference-affiliated teams than other teams. Since the *DIST* variables are not dummies, results from two separate tests are shown: one examines how pollster accuracy changes when the teams move from 750 miles closer than their respective mean values (shown in Table 3) to their mean values, and the other shows accuracy when they move from their respective mean values to 750 miles further away.

When including games involving unranked teams in the sample and omitting *RANK* controls (Column 1), there is no evidence that pollsters more accurately order teams in closer geographic proximity. Adding *RANK* controls (Column 2) changes virtually nothing. Columns 3-4 repeat Columns 1-2 but omit unranked teams. In Column 3, the *STATE* coefficients sum to

1.162, and the sum is significantly positive at the 13% level, providing some weak evidence that pollsters are more likely to correctly order own-state teams when both are ranked. The sum of the *CONFERENCE* coefficients only has a *p*-value of 0.261. When moving from 750 miles closer than the mean to the mean, pollster accuracy improves, and the effect is significant at the 14% level. This weakly suggests that pollsters order more-nearby teams *less* accurately than other teams, *ceteris paribus*, perhaps because they underrate them too frequently. Results including *RANK* controls (Column 4) are very similar.[21] Columns 5-8 repeat Columns 1-4 but perform linear probability models instead of logit estimations. Results are, again, similar but weaker. Coefficients in linear probability estimations can be interpreted as marginal effects.

V. Conclusion

This paper uses poll data and game outcome data from the 2009-2010 college basketball season to determine whether AP Poll voters exhibit geographical discrimination in their rankings, and to further discern whether such discrimination constitutes taste-based or information-based discrimination. Results show that pollsters vote more-nearby teams (which are fewer miles away) to significantly higher rankings, especially at the bottom of their polls. In the bottom half of polls, own-state teams are ranked to higher ranks, but at the top of polls conference-affiliated teams are actually voted to lower ranks. More-nearby teams, own-state teams, and conference-affiliated teams are more frequently ranked in the Top 25 than other teams.

Estimations on game-outcome data show no evidence that pollsters overrate, i.e. engage in taste-based discrimination favoring, teams in close geographic proximity. When pollsters place an own-state or conference-affiliated team to a higher ranking, there is no significant loss in accuracy, and when they place a more-nearby team to a higher ranking, they increase their accuracy. When pollsters vote own-state teams and conference-affiliated teams to worse rankings, they improve their accuracy, correctly inferring that these teams are of worse quality. But when they vote more-nearby teams to worse rankings, they reduce their accuracy. This last result suggests voters underrate more-nearby teams, perhaps because they are hesitant to be perceived as engaging in taste-based discrimination. The effect is especially strong at the top of polls. Other estimations provide some weak evidence that pollsters are more likely to correctly order own-state teams than out-of-state teams, but also that they order more-nearby teams less accurately that teams that are the mean distance away.

The research in this paper could be extended in several directions. The relationship between geography and pollster behavior could be studied in a more dynamic framework. Pollster responses to performance—such as wins, losses, margin of victory, etc.—could be examined to see if nearby pollsters react differently than farther-away pollsters. Furthermore, the relationship between pollsters and fans could be measured to see if pollsters are influenced by customer discrimination [Holzer and Ihlanfeldt 1998; Burdekin et al 2005]. With polls and pollster contact information being publicly available, and with more media members having websites, blogs, or twitter accounts where vocal fans often post comments,[22] pollsters may feel pressure to vote local teams to higher ranks. Perhaps an interesting future topic of research could entail whether pollsters who have more interaction with fans have higher rates of taste-based geographic discrimination.

Acknowledgements

The author wishes to especially thank Daniel F. Stone for comments on earlier drafts. He also wishes to thank audiences at Washington State University and anonymous referees for comments.

---

[1] Aigner and Cain [1977] insist that mistake-based discrimination differs from information-based discrimination because systemic inaccuracies about certain groups cannot survive in a competitive marketplace, but Romer [2006] finds that systemic failures to optimize do indeed exist in competition.

[2] Other sources of discrimination in polls have been found as well: Mirabile and Witte [2010] find that coaches voting in the 2004-2008 *USA Today* Polls ranked opponents of their alma mater to higher rankings than other teams.

[3] There were 65 pollsters at any point in the season, but 66 pollsters overall. Patrick Stevens of *The Washington Times* was dismissed as a pollster after the *Times* shut down its sports section. This occurred after the 9th poll of the season had been released. He was replaced by John McNamara of *The Capital* in Annapolis, Maryland.

[4] The *USA Today* Coaches Poll determines one-third of the Bowl Championship Series (BCS) standings, which determines who plays in the national championship game. See http://www.bcsknowhow.com/bcs-formula. Dummett [1998] shows how pollsters can manipulate Borda count polls.

[5] The National Collegiate Athletics Association (NCAA) postseason tournament had 65 teams in the 2009-2010 season. Since 2010-2011 it has had 68 teams [Prisbell and Yanda 2010]. Coleman, DuMond, and Lynch [2010] find that the Sagarin Computer Poll ranking can affect the seeding of NCAA tournament teams, but find no evidence it affects whether a team makes the tournament or not.

[6] This difference between college football and college basketball may be mitigated if college football pollsters are subject to more oversight or public scrutiny than college basketball pollsters.

[7] College basketball polls also appear to have been the subject of much less research than college football polls. A large, multi-disciplinary bibliography of research regarding college football rankings can be found at http://homepages.cae.wisc.edu/~dwilson/rsfc/rate/biblio.html.

[8] The three pollsters situated in Washington, DC are included as being in Virginia and Maryland as well as Washington, DC.
[9] Distances are measured via the zip codes of universities and pollsters' employers. They account for curvature of the earth.
[10] Kate Hairopolous of *The Dallas Morning News* was the only female of the 66 pollsters in 2009-2010.
[11] A t subscript could feasibly be included in the subscripts for X_{ip} and σ_p because the population of pollsters changed over the course of the season (see Footnote 1).
[12] There are 130 fewer observations in Table 2 Columns 1-2 than in Table 1 Panel A because twice during the season all 65 first-place votes went to the same team (Kansas). These 130 observations were removed because the team-week fixed effect coefficient was beyond the upper-level censor and not identified, and the team's inclusion provided no extra information to the estimation.
[13] This contrasts with Coleman et al [2010], who found distance to be insignificantly related to votes in the AP College Football Poll.
[14] Wooldridge [2002] shows how to create a predicted value from a one-sided Tobit. Creating a predicted value from a two-sided Tobit is an extension of the model.
[15] In the bottom-half sample estimation, B_{ipt} in Equation (3) is upper-bounded at 13.
[16] Coleman [2005] also discusses "retrodictive accuracy," by which poll accuracy is measured vis-à-vis how well it reflects completed team performance over the course of a season.
[17] Having separate controls for the better-ranked and worse-ranked teams is similar to gravity models where migration can vary with the characteristics of both origin and destination cities [e.g. Fields 1976].
[18] Among games included in the sample, the mean value of *WORSE_RANK* differs slightly between Panel A and Panel B. This is because not every Panel A game involving two ranked teams is observed in Panel B. In some games between, for example, Team A and Team B, Team A is ranked above Team B in all observations where both are ranked, but Team B is ranked in some polls where Team A is unranked. Thus, when removing unranked observations from the dataset, the remaining observations are perfectly classified and are omitted from the estimations.
[19] The rank coefficients themselves (not shown) confirm that rankings are more accurate at the top of the rankings than at the bottom of the rankings.
[20] In Column 2, *CONFERENCE_MATCH_BETTER*BETTER_RANK* is positive at the 13% level.
[21] Estimations were also performed where own-state teams are counted as conference-affiliated teams, and the controls for *STATE_MATCH_BETTER*, *STATE_MATCH_WORSE*, and their interaction were removed from the sample. The sum of the three *CONFERENCE* coefficients was significant at the 15% level when omitting games involving unranked teams.
[22] For examples, see John Feinstein's website (http://www.feinsteinonthebrink.com) or John Wilner's blog (http://blogs.mercurynews.com/collegesports).

References

Ai, Chunrong; Norton, Edward C. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80:1, pp. 123-129.

Aigner, Dennis J.; Cain, Glen G. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review 30:2*, pp. 175-187.

Altonji, Joseph G.; Pierret, Charles R. 2001. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics 116:1*, pp. 313-350.

Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: The University of Chicago Press.

Bertrand, Marianne; Chugh, Dolly; Mullainathan, Sendhil. 2005. "Implicit Discrimination." *American Economic Review* 95:2, pp. 94-98.

Burdekin, Richard C. K.; Hossfeld, Richard T.; Smith, Janet Kiholm. 2005. "Are NBA Fans Becoming Indifferent to Race? Evidence From the 1990s." *Journal of Sports Economics 6:2*, pp. 144-159.

Campbell, Noel D.; Rogers, Tammy M.; Finney, R. Zachary. 2007. "Evidence of Television Exposure Effects in AP Top 25 College Football Rankings." *Journal of Sports Economics 8:4*, pp. 425-434.

Coleman, B. Jay. 2005. "Minimizing Game Score Violations in College Football Rankings." *Interfaces 35:6*, pp. 483-496.

Coleman, B. Jay; Gallo, Andres; Mason, Paul M.; Steagall Jeffrey W. 2010. "Voter Bias in the Associated Press College Football Poll." *Journal of Sports Economics 11:4,* pp. 397-417.

Dummett, Michael. 1998. "The Borda Count and Agenda Manipulation." *Social Choice and Welfare* 15:2, pp. 289-296.

Ehrenberg, Ronald G.; Smith, Robert M. 2009. *Modern Labor Economics: Theory and Public Policy, 10th Edition*. San Francisco: Pearson Education.

Fowdur, Lona; Kadiyali, Vrinda; Prince, Jeffrey. 2012. "Racial Bias in Expert Quality Assessment: A Study of Newspaper Movie Reviews." *Journal of Economic Behavior and Organization* 84:1, pp. 292-307.

Gneezy, Uri; List, John; Price, Michael K. 2012. "Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments." NBER Working Paper No. 17855.

Hausman, Jerry A.; Rudd, Paul A. 1987. "Specifying and testing econometric models for rank-ordered data." *Journal of Econometrics 34:1-2*, pp. 83-104.

Holzer, Harry J.; Ihlanfeldt, Keith R. 1998."Customer Discrimination and Employment Outcomes for Minority Workers." *The Quarterly Journal of Economics 113:3*, pp. 835-867.

Holzer, Harry J.; Raphael, Stephen; Stoll, Michael A. 2006. "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers." *Journal of Law and Economics 44:2*, pp. 451-480.

LeSage, James; Pace, Robert Kelly. 2009. *Introduction to Spatial Econometrics*. London: Chapman & Hall.

Levitt, Steven D. 2004. "Testing Theories of Discrimination: Evidence from Weakest Link." *Journal of Law and Economics 47:2*, pp. 431-453.

List, John A. 2004. "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field." *The Quarterly Journal of Economics* 119:1, pp. 49-89.

Logan, Trevon D. 2007. "Whoa, Nellie! Empirical Tests of College Football's Conventional Wisdom." NBER Working Paper 13596.

Mandel, Stewart. 2005. "The *Real* BCS Controversy: Want Evidence of Bias? Just Look at Coaches' Votes." *SI.com.*

Mirabile, McDonald Paul; Witte, Mark David. 2010. "Not So Fast, My Friend: Biases in College Football Polls." *Journal of Sports Economics 11:4,* pp. 443-455.

Morris, Stephen. "Political Correctness." 2001. *Journal of Political Economy* 109:2, pp. 231-265.

Nutting, Andrew W. 2011. "…And After That, Who Knows?: Detailing the Marginal Accuracy of Weekly College Football Polls." *Journal of Quantitative Analysis in Sports 7:3*, Article 7.

Parsons, Christopher A.; Sulaeman, Johan; Yates, Michael C.; Hamermesh, Daniel S. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." American Economic Review 101:4, pp. 1410-1435.

Price, Joseph P.; Wolfers, Justin. 2010. "Racial Discrimination Among NBA Referees." *Quarterly Journal of Economics 12*5:4, pp. 1859-1877.

Prisbell, Eric; Yanda, Steve. 2010. "NCAA men's tourney will expand to 68; Three teams added in 14-year, $10.8 billion deal reached with CBS, Turner." The Washington Post. April 23, Page D-1.

Romer, David. 2006. "Do Firms Maximize? Evidence from Professional Football." *Journal of Political Economy 114:2*, pp. 340-365.

Ross, Justin M.; Larson, Sarah E.; Wall, Chad. 2012. "Are Surveys of Experts Unbiased? Evidence from College Football Rankings." *Contemporary Economic Policy 30:4*, pp. 502-22.

Shayo, Moses; Zussman, Asaf. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *Quarterly Journal of Economics 126:3*, pp. 1447-1484.

Steinberg, Dan. 2007. "The AP's AU Voter: John Feinstein." *Washington Post D.C. Sports Bog*. December 28.

Stone, Daniel F. 2013. "Testing Bayesian Updating with the Associated Press Top 25." *Economic Inquiry 51:2*, pp. 1457-74.

Stone, Daniel F.; Zafar, Basit. "Do We Follow Others When We Should Outside the Lab? Evidence from the AP Top 25." Forthcoming, Journal of Risk and Uncertainty.

Wolfers, Justin. 2006. "Diagnosing Discrimination: Stock Returns and CEO Gender." *Journal of the European Economic Association* 4:2/3, pp. 531-541.

Wolfers, Justin; Kumar, Alok. 2008. "Do Male Analysts, Mutual Fund Managers, and Individual Investors Discriminate Against Female CEOs?" Working Paper, University of Pennsylvania.

Table 1
Summary Statistics of Pollster/Team Observations

	Panel A: All Observations Obs = 59,150				Panel B: Ranked Teams Only Obs = 30,875			
	Mean	**St. Dev**	**Min**	**Max**	**Mean**	**St. Dev**	**Min**	**Max**
RANKED	0.52	0.50	0	1	1	-	1	1
POINTS	6.8	8.3	0	25	13.0	7.2	1	25
STATE_MATCH	0.031	0.17	0	1	0.034	0.18	0	1
CONFERENCE_MATCH	0.206	0.40	0	1	0.219	0.41	0	1
DIST	0.896	0.615	0	2.736	0.863	0.607	0	2.736

Table 2
Estimations of Equation (1)
* significant at 10%; ** significant at 5%; *** significant at 1%
Standard errors in brackets

	1	**2**	**3**	**4**	**5**
STATE_MATCH	1.064***	0.709***	0.438***	1.182***	-0.012
	[0.120]	[0.133]	[0.056]	[0.228]	[0.096]
CONFERENCE_MATCH	0.295***	0.074	0.097***	0.092	-0.121***
	[0.053]	[0.059]	[0.025]	[0.100]	[0.046]
DIST		-0.418***	-0.150***	-0.661***	-0.172***
		[0.055]	[0.023]	[0.087]	[0.051]
Estimation	Tobit	Tobit	Logit	Tobit	Tobit
Rankings in Estimation	1-25	1-25	Ranked/ Unranked	13-25	1-13
Observations	59,020	59,020	43,355	50,700	29,770
Log Pseudo-likelihood	-89,523.9	-89,491.3	-13,937.4	-56,390.8	-35,854.7

Table 3
Summary Statistics of Pollster/Game Observations

	PANEL A: INCLUDING GAMES WITH UNRANKED TEAMS							
	In Sample Obs = 5,484 (120 games)				Not in Sample Obs = 45,223 (1,261 games)			
	Mean	**St. Dev**	**Min**	**Max**	**Mean**	**St. Dev**	**Min**	**Max**
WIN	0.535	0.499	0	1	0.811	0.392	0	1
BETTER_RANK	14.8	6.4	1	25	12.1	7.3	1	25
*WORSE_RANK**	16.1	6.2	3	25	16.4	5.4	5	25
WORSE_UNRANKED	0.473	0.499	0	1	0.926	0.261	0	1
STATE_MATCH_BETTER	0.037	0.188	0	1	0.034	0.182	0	1
CONFERENCE_MATCH_BETTER	0.209	0.407	0	1	0.217	0.412	0	1
DIST_BETTER	0.838	0.598	0	2.736	0.866	0.607	0	2.736
STATE_MATCH_WORSE	0.030	0.171	0	1	0.031	0.174	0	1
CONFERENCE_MATCH_WORSE	0.210	0.408	0	1	0.193	0.395	0	1
DIST_WORSE	0.875	0.605	0	2.736	0.921	0.635	0	5.076

	PANEL B: OMITTING GAMES WITH UNRANKED TEAMS							
	In Sample Obs = 2,847 (76 games)				Not in Sample Obs = 3,375 (140 games)			
	Mean	**St. Dev**	**Min**	**Max**	**Mean**	**St. Dev**	**Min**	**Max**
WIN	0.596	0.491	0	1	0.688	0.463	0	1
BETTER_RANK	10.8	5.2	1	24	4.8	3.8	1	24
WORSE_RANK	16.0	6.1	3	25	16.5	5.5	5	25
WORSE_UNRANKED								
STATE_MATCH_BETTER	0.034	0.181	0	1	0.038	0.191	0	1
CONFERENCE_MATCH_BETTER	0.224	0.417	0	1	0.226	0.418	0	1
DIST_BETTER	0.788	0.591	0.001	2.674	0.799	0.571	0.002	2.674
STATE_MATCH_WORSE	0.033	0.178	0	1	0.037	0.189	0	1
CONFERENCE_MATCH_WORSE	0.230	0.421	0	1	0.228	0.420	0	1
DIST_WORSE	0.832	0.605	0.001	2.627	0.800	0.572	0.001	2.627

* = Panel A *WORSE_RANK* values only for games between ranked teams

Table 4
Interpretation of Coefficients in Equation (4)

	1 Better Information	2 Worse Information	3 Overrated	4 Underrated
STATE				
STATE_MATCH_BETTER	+	-	-	0 or +
STATE_MATCH_WORSE	+	-	0 or +	-
CONFERENCE				
CONFERENCE_MATCH_BETTER	+	-	-	0 or +
CONFERENCE_MATCH_WORSE	+	-	0 or +	-
DISTANCE				
DIST_BETTER	-	+	+	0 or -
DIST_WORSE	-	+	0 or -	+

Table 5
Estimations of Equation (2)
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors in brackets

	1	2	3	4
	All Games		Worse Team Ranked	
STATE_MATCH_BETTER	0.270	0.308	0.449	0.518
	[0.258]	[0.255]	[0.390]	[0.391]
STATE_MATCH_WORSE	0.087	0.122	0.061	0.056
	[0.300]	[0.301]	[0.356]	[0.367]
CONFERENCE_MATCH_BETTER	0.127	0.167	0.161	0.223
	[0.111]	[0.112]	[0.188]	[0.191]
CONFERENCE_MATCH_WORSE	0.041	0.039	0.003	-0.018
	[0.142]	[0.139]	[0.152]	[0.147]
RANK controls	No	Yes	No	Yes
Observations	5,484	5,484	2,847	2,847
Log Pseudo-likelihood	-1733.9	-1719.8	-1060.2	-1043.8

Table 6
Estimations of Equation (4)
* significant at 10%; ** significant at 5%; *** significant at 1%
Standard errors in brackets

	1	2	3	4	5	6	7	8
	Conference Games				**Non-Conference Games**			
	All Games		**Worse Team Ranked**		**All Games**		**Worse Team Ranked**	
STATE_MATCH_BETTER	0.611*	0.608*	0.640	0.696	-0.193	-0.131	0.167	0.328
	[0.320]	[0.315]	[0.445]	[0.440]	[0.312]	[0.311]	[0.496]	[0.493]
STATE_MATCH_WORSE	0.078	0.125	0.003	-0.027	0.041	0.045	0.017	-0.028
	[0.440]	[0.443]	[0.419]	[0.435]	[0.370]	[0.366]	[0.502]	[0.508]
CONFERENCE_MATCH_BETTER	0.066	0.097	-0.013	0.003	0.269	0.308*	0.401	0.488*
	[0.192]	[0.192]	[0.232]	[0.224]	[0.183]	[0.186]	[0.268]	[0.274]
CONFERENCE_MATCH_WORSE					0.098	0.111	0.212	0.233
					[0.182]	[0.177]	[0.221]	[0.223]
RANK controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,896	2,896	1,392	1,392	2,588	2,588	1,455	1,455
Log Pseudo-likelihood	-892.9	-886.6	-483.5	-476.1	-808.5	-796.1	-540.3	-527.5

Table 7
Estimations of Equation (4)
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors in brackets

	1	2	3	4
	All Games		**Worse Team Ranked**	
STATE_MATCH_BETTER	-0.024	0.009	0.038	0.114
	[0.297]	[0.295]	[0.426]	[0.429]
STATE_MATCH_WORSE	0.469	0.485	0.705*	0.690*
	[0.300]	[0.302]	[0.370]	[0.378]
CONFERENCE_MATCH_BETTER	-0.076	-0.039	-0.134	-0.070
	[0.145]	[0.145]	[0.228]	[0.229]
CONFERENCE_MATCH_WORSE	0.273*	0.260*	0.394**	0.366**
	[0.154]	[0.151]	[0.181]	[0.176]
DIST_BETTER	-0.435***	-0.438***	-0.726***	-0.722***
	[0.166]	[0.165]	[0.220]	[0.216]
DIST_WORSE	0.504***	0.486***	0.924***	0.925***
	[0.127]	[0.129]	[0.200]	[0.206]
RANK controls	No	Yes	No	Yes
Observations	5,484	5,484	2,847	2,847
Log Pseudo-likelihood	-1723.3	-1709.7	-1038.8	-1022.8

Table 8
Estimations of Equation (4)
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors in brackets

	1	2	3	4	5	6	7	8
	Conference Games				Non-Conference Games			
	All Games		Worse Team Ranked		All Games		Worse Team Ranked	
STATE_MATCH_BETTER	0.420	0.406	0.492	0.538	-0.400	-0.348	-0.130	0.041
	[0.403]	[0.400]	[0.565]	[0.564]	[0.354]	[0.349]	[0.524]	[0.519]
STATE_MATCH_WORSE	0.374	0.412	0.347	0.316	0.420	0.396	0.780	0.742
	[0.451]	[0.454]	[0.436]	[0.450]	[0.381]	[0.378]	[0.544]	[0.570]
CONFERENCE_MATCH_BETTER	0.095	0.110	0.149	0.151	0.135	0.162	0.194	0.274
	[0.235]	[0.233]	[0.324]	[0.314]	[0.222]	[0.223]	[0.328]	[0.330]
CONFERENCE_MATCH_WORSE					0.313*	0.313*	0.611**	0.627**
					[0.186]	[0.187]	[0.238]	[0.250]
DIST_BETTER	-1.189***	-1.186***	-2.136***	-2.167***	-0.319	-0.330	-0.521*	-0.529*
	[0.353]	[0.362]	[0.508]	[0.529]	[0.209]	[0.205]	[0.283]	[0.274]
DIST_WORSE	1.266***	1.246***	2.204***	2.221***	0.456***	0.432***	0.969***	0.979***
	[0.322]	[0.333]	[0.503]	[0.515]	[0.148]	[0.152]	[0.220]	[0.234]
RANK controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,896	2,896	1,392	1,392	2,588	2,588	1,455	1,455
Log Pseudo-likelihood	-892.9	-886.6	-483.5	-476.1	-808.5	-796.1	-540.3	-527.5

Table 9
Estimations of Equation (4) with ordinal rank interactions
No Games Featuring Unranked Teams
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors in brackets

	1	**2**	**3**	**4**	**5**	**6**
STATE_MATCH_BETTER	0.466	-0.106		0.027	-0.044	
	[0.923]	[0.986]		[0.088]	[0.094]	
*STATE_MATCH_BETTER*BETTER_RANK*	0.003	0.015		0.003	0.003	
	[0.079]	[0.088]		[0.008]	[0.009]	
STATE_MATCH_WORSE	-0.531	0.432		-0.095	0.003	
	[0.989]	[0.901]		[0.114]	[0.114]	
*STATE_MATCH_WORSE*WORSE_RANK*	0.031	0.013		0.005	0.004	
	[0.049]	[0.048]		[0.006]	[0.006]	
CONFERENCE_MATCH_BETTER	-0.245	-0.720*	-0.711*	-0.043	-0.090**	-0.086*
	[0.316]	[0.390]	[0.391]	[0.038]	[0.044]	[0.045]
*CONFERENCE_MATCH_BETTER*BETTER_RANK*	0.041	0.055	0.057*	0.006	0.006	0.006
	[0.032]	[0.036]	[0.034]	[0.004]	[0.005]	[0.004]
CONFERENCE_MATCH_WORSE	0.077	0.702	0.673	0.015	0.086	0.087
	[0.446]	[0.519]	[0.491]	[0.054]	[0.061]	[0.057]
*CONFERENCE_MATCH_WORSE*WORSE_RANK*	-0.008	-0.022	-0.026	-0.002	-0.003	-0.003
	[0.029]	[0.033]	[0.031]	[0.004]	[0.004]	[0.004]
DIST_BETTER		-1.831***	-1.799***		-0.175***	-0.168***
		[0.498]	[0.479]		[0.057]	[0.053]
*DIST_BETTER*BETTER_RANK*		0.080**	0.076**		0.005	0.004
		[0.035]	[0.033]		[0.004]	[0.004]
DIST_WORSE		2.203***	2.126***		0.211***	0.205***
		[0.612]	[0.598]		[0.059]	[0.058]
*DIST_WORSE*WORSE_RANK*		-0.071**	-0.073***		-0.005*	-0.005*
		[0.028]	[0.027]		[0.003]	[0.003]
Estimation	Logit	Logit	Logit	LPM	LPM	LPM
RANK controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,847	2,847	2,847	2,847	2,847	2,847
Log Pseudo-likelihood	-1042.5	-1016.8	-1018.7			
R^2				0.516	0.526	0.525

Table 10
Estimations of Equation (4) with geographic interaction terms
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors in brackets

	1	2	3	4	5	6	7	8
	All Games		Both Teams Ranked		All Games		Both Teams Ranked	
STATE_MATCH_BETTER	-0.053	-0.025	0.043	0.112	-0.012	-0.011	-0.015	-0.009
	[0.316]	[0.310]	[0.458]	[0.455]	[0.032]	[0.032]	[0.048]	[0.047]
STATE_MATCH_WORSE	0.445	0.456	0.665	0.645	0.048	0.048	0.083	0.082
	[0.316]	[0.317]	[0.426]	[0.430]	[0.033]	[0.033]	[0.051]	[0.051]
STATE_INTERACT	-0.118	-0.083	0.454	0.477	-0.005	0.002	0.040	0.038
	[0.736]	[0.749]	[0.850]	[0.875]	[0.083]	[0.085]	[0.107]	[0.109]
CONFERENCE_MATCH_BETTER	-0.008	0.033	0.147	0.230	-0.009	-0.006	-0.005	0.001
	[0.203]	[0.204]	[0.325]	[0.325]	[0.019]	[0.019]	[0.035]	[0.035]
CONFERENCE_MATCH_WORSE	0.337	0.329	0.650**	0.640**	0.033	0.033	0.070**	0.069**
	[0.215]	[0.214]	[0.269]	[0.272]	[0.020]	[0.020]	[0.030]	[0.030]
CONFERENCE_INTERACT	-0.169	-0.181	-0.516	-0.556	-0.013	-0.013	-0.047	-0.049
	[0.295]	[0.302]	[0.427]	[0.432]	[0.029]	[0.029]	[0.049]	[0.048]
DIST_BETTER	-0.477*	-0.481*	-0.361	-0.353	-0.046*	-0.047*	-0.076	-0.076
	[0.256]	[0.260]	[0.362]	[0.368]	[0.026]	[0.026]	[0.047]	[0.047]
DIST_WORSE	0.459*	0.441*	1.306***	1.312***	0.055**	0.053*	0.163***	0.163***
	[0.254]	[0.256]	[0.386]	[0.391]	[0.027]	[0.027]	[0.044]	[0.045]
DIST_INTERACT	0.041	0.041	-0.326	-0.328	-0.001	0.000	-0.034	-0.035
	[0.200]	[0.201]	[0.285]	[0.291]	[0.020]	[0.020]	[0.034]	[0.035]
Estimation	Logit	Logit	Logit	Logit	LPM	LPM	LPM	LPM
RANK controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	5,484	5,484	2,847	2,847	5,484	5,484	2,847	2,847
Log Pseudo-likelihood	-1723.1	-1709.5	-1036.8	-1020.6				
R^2					0.612	0.614	0.520	0.525
Sums of coefficients with p-values								
STATE	0.274	0.348	1.162	1.234	0.031	0.039	0.108	0.111
	[0.687]	[0.623]	[0.123]	[0.125]	[0.676]	[0.625]	[0.263]	[0.281]
CONFERENCE	0.160	0.181	0.281	0.314	0.011	0.014	0.018	0.021
	[0.389]	[0.330]	[0.261]	[0.215]	[0.533]	[0.462]	[0.516]	[0.459]
DIST								
From 750 Miles Closer than Mean	0.016	0.000	0.494	0.503	0.006	0.004	0.043	0.042
to Mean	[0.940]	[1.000]	[0.131]	[0.136]	[0.722]	[0.853]	[0.271]	[0.285]
From Mean to 750 Miles Further	0.062	0.046	0.107	0.113	0.005	0.004	0.005	0.004
Than Mean	[0.677]	[0.761]	[0.491]	[0.493]	[0.736]	[0.802]	[0.276]	[0.892]

Figure 1: State Effects

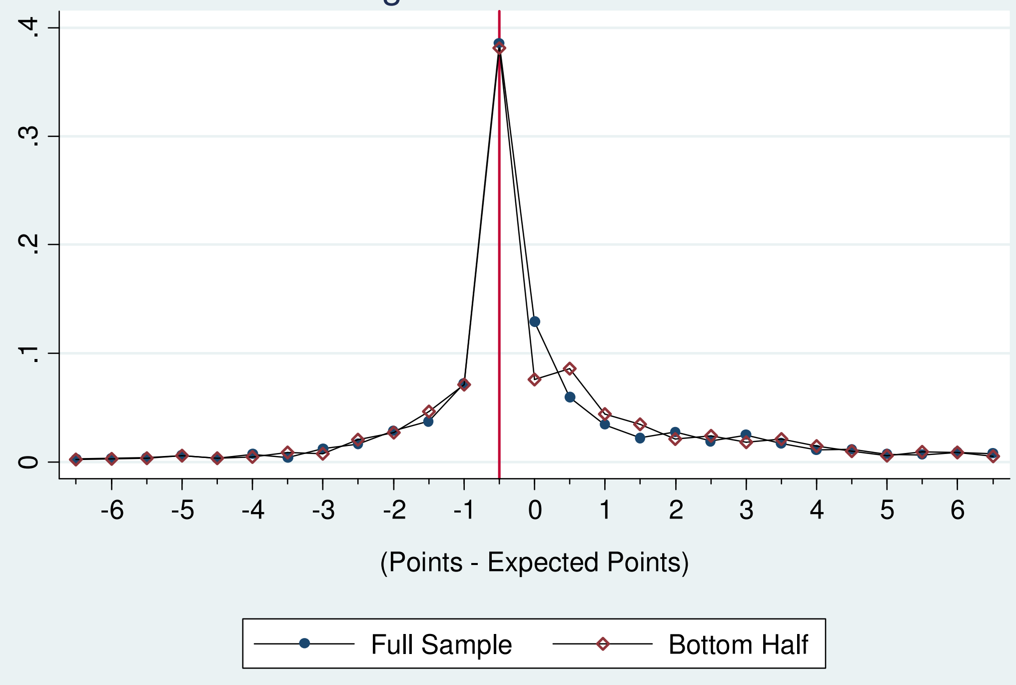


Figure 2: Distance Effects, Full Sample

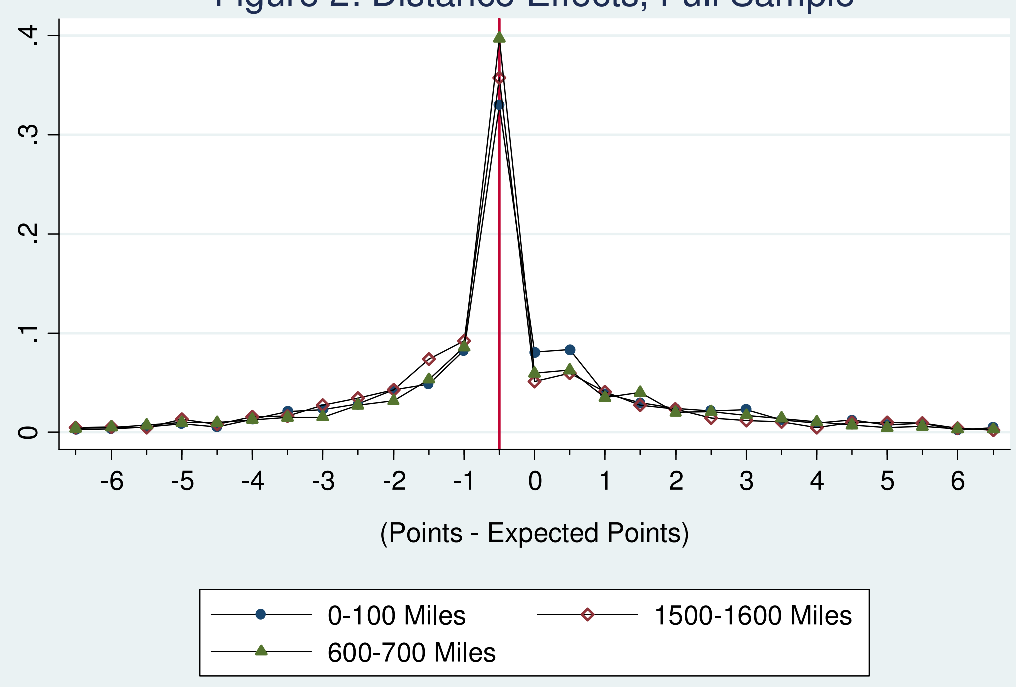


Figure 3: Distance Effects, Bottom Half

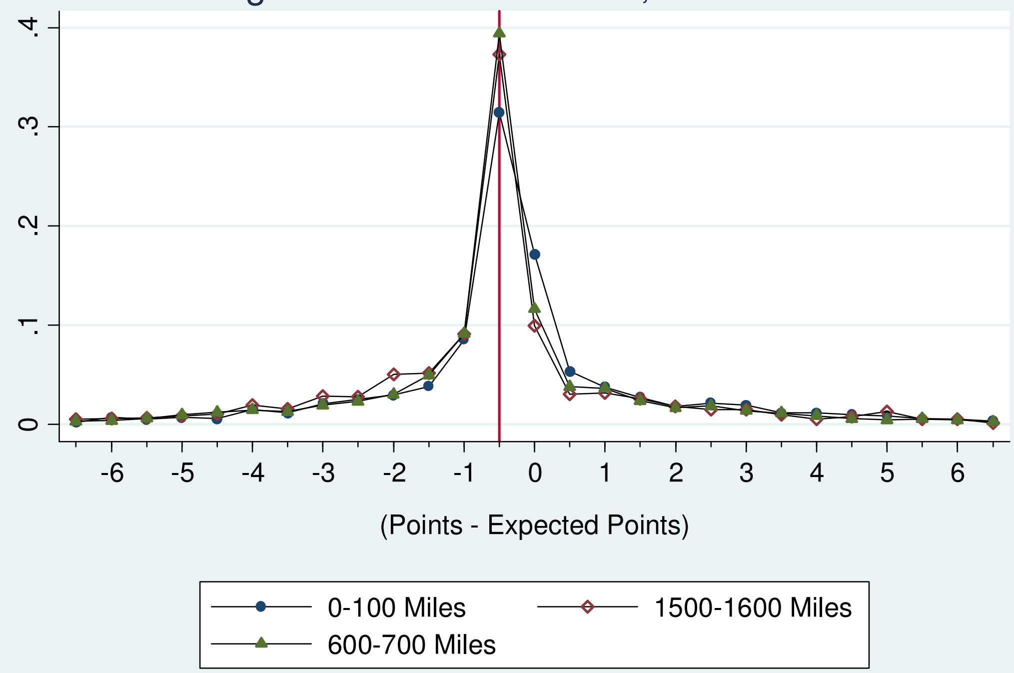
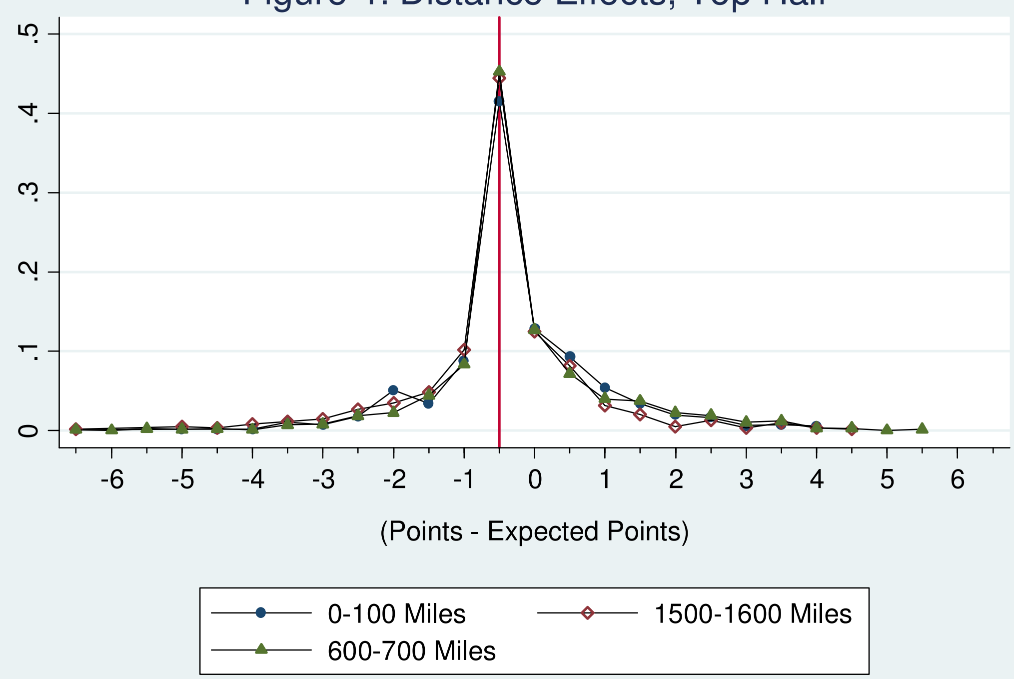


Figure 4: Distance Effects, Top Half



Appendix Table 1
List of 2009-10 AP Basketball Poll Voters

	Pollster	Employer	Location
1	Elton Alexander	The Cleveland Plain Dealer	Cleveland, OH
2	Al Balderas	The Orange County Register	Santa Ana, CA
3	Mark Berman	The Roanoke Times	Roanoke, VA
4	Jack Bogaczyk	Charleston Daily Mail	Charleston, WV
5	John Bohnenkamp	The Hawk Eye	Burlington, IA
6	Dave Borges	New Haven Register	New Haven, CT
7	Rich Bozich	The Courier-Journal	Louisville, KY
8	Steven Bradley	The Journal	Seneca, SC
9	David Brandt	The Clarion-Ledger	Jackson, MS
10	Kevin Brockway	The Gainesville Sun	Gainesville, FL
11	Roger Clarkson	The Athens Banner-Herald	Athens, GA
12	Bill Cole	Winston-Salem Journal	Winston-Salem, NC
13	Cory Curtis	WKRN	Nashville, TN
14	Seth Davis	Sports Illustrated	New York, NY
15	Steve DeShazo	The Free Lance-Star	Fredericksburg, VA
16	John Feinstein	National Public Radio	Washington, DC
17	Jason Franchuk	Provo Daily Herald	Provo, UT
18	Pete Gilbert	WBAL	Baltimore, MD
19	Charles Goldberg	The Birmingham News	Birmingham, AL
20	Jeff Goodman	Foxsports.com	West Stockbridge, MA
21	Cormac Gordon	Staten Island Advance	Staten Island, NY
22	Ed Graney	Las Vegas Review-Journal	Las Vegas, NV
23	Vahe Gregorian	St. Louis Post-Dispatch	St. Louis, MO
24	Jason Groves	Las Cruces Sun-News	Las Cruces, NM
25	Kate Hairopoulos	Dallas Morning News	Dallas, TX
26	Tim Hall	WCMC-FM	Raleigh, NC
27	Doug Haller	The Arizona Republic	Phoenix, AZ
28	Bob Holt	Arkansas Democrat-Gazette	Little Rock, AR
29	Gary Horowitz	Salem Statesman Journal	Salem, OR
30	Terry Hutchens	The Indianapolis Star	Indianapolis, IN
31	Pete Iorizzo	Times Union	Albany, NY
32	Nick Jezierny	The Idaho Statesman	Boise, ID
33	Scott Johnson	The Daily Herald	Everett, WA
34	Joey Johnston	The Tampa Tribune	Tampa, FL
35	Dave Jones	The Patriot-News	Harrisburg, PA
36	Joe Juliano	The Philadelphia Inquirer	Philadelphia, PA
37	Tom Keegan	Lawrence Journal World	Lawrence, KS
38	Paul Klee	Champaign News-Gazette	Champaign, IL
39	Gary Laney	The Advocate	Baton Rouge, LA
40	Chris Lang	The News & Advance	Lynchburg, VA
41	Dave Mackall	Pittsburgh Tribune-Review	Pittsburgh, PA
42	Scott Mansch	Great Falls Tribune	Great Falls, MT
43	Jeffrey Martin	Houston Chronicle	Houston, TX
44	Mark McCarter	The Huntsville Times	Huntsville, AL
45	Matt McCoy	WTVN-AM	Columbus, OH
46	Kevin McNamara	The Providence Journal	Providence, RI

Appendix Table 1 (continued)
List of 2009-10 AP Basketball Poll Voters

47 John McNamara	The Capital	Annapolis, MD
48 Ron Morris	The State	Columbia, SC
49 Joshua Parrott	The Daily Advertiser	Lafayette, LA
50 Lamond Pope	The Journal Gazette	Fort Wayne, IN
51 Pat Ridgell	Daily Times-Call	Longmont, CO
52 John Rohde	The Oklahoman	Oklahoma City, OK
53 Brian Rosenthal	Lincoln Journal Star	Lincoln, NE
54 Todd Rosiak	Milwaukee Journal-Sentinel	Milwaukee, WI
55 Michael Rothstein	annarbor.com	Ann Arbor, MI
56 Keith Sargeant	Gannett NJ Newspapers	Neptune, NJ
57 Patrick Stevens	Washington Times	Washington, DC
58 Craig Stouffer	The Washington Examiner	Washington, DC
59 Bob Sutton	Burlington Times-News	Burlington, NC
60 Dick Vitale	ABC-ESPN	New York, NY
61 Dick Weiss	Daily News	New York, NY
62 John Werner	Waco Tribune-Herald	Waco, TX
63 Lindsey Willhite	The Daily Herald	Arlington Heights, IL
64 Jon Wilner	San Jose Mercury News	San Jose, CA
65 Scott Wolf	Los Angeles Daily News	Los Angeles, CA
66 Dan Wolken	The Commerical Appeal	Memphis, TN

Appendix Table 2
Estimations of Equation (1), Ordered Probits
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors in brackets
Marginal effects in braces

	1	**2**	**3**	**4**	**5**
STATE_MATCH	0.275***	0.184***		0.216***	0.035
	[0.033]	[0.036]		[0.045]	[0.050]
	{0.490}	{0.324}		{0.333}	{0.034}
CONFERENCE_MATCH	0.072***	0.015		0.012	-0.049**
	[0.015]	[0.017]		[0.019]	[0.024]
	{0.125}	{0.026}		{0.018}	{-0.047}
DIST		-0.107***		-0.129***	-0.076***
		[0.015]		[0.017]	[0.025]
		{-0.184}		{-0.194}	{-0.073}
Rankings in Estimation	1-25	1-25		13-25	1-13
Observations	59,020	59,020		50,700	29,770
Log Pseudo-likelihood	-80,682.0	-80,654.8		-54,802.3	-32,627.8