International Journal of Sport Finance, 2013, **8**, 175-191, © 2013 West Virginia University

Failing to Cover: Point Shaving or Statistical Abnormality?

George Diemer[1] and Michael A. Leeds[2]

[1]Chestnut Hill College
[2]Temple University

George Diemer is an assistant professor of business at Chestnut Hill College. His current research interests include gambling market for sporting events as it translates to the financial market, corruption in sporting events, and industrial organization.

Michael A. Leeds is a professor of economics and director of graduate studies in economics at Temple University. His current research focuses on business models of Japanese baseball and on differences in how men and women respond to economic contests.

Abstract

The possibility that coaches, players, or referees might be involved in point shaving has been a subject of debate since Wolfers's (2006) controversial finding that favorites in NCAA college basketball games fail to cover point spreads with disturbing frequency. We reconcile Wolfers's finding with evidence provided by Borghesi (2008), Borghesi and Dare (2009), and others that heavy favorites are not, on average, less likely to cover the point spread. We find that the distribution of game outcomes is bimodal, with one peak on one side of the "no corruption" outcome and one peak on the other side. This finding is consistent with point shaving by favorites, who lose by too little, and underdogs who lose by too much. On average, however, the outcome is consistent with the no-point shaving hypothesis. We compare regular-season and post-season results and find that this effect disappears in the more closely observed NCAA tournament games.

Keywords: corruption, gambling, point shaving, NCAA, basketball

Introduction

Gambling and game-tampering have accompanied the development of team sports in America. According to Seymour (1960), the Mutual Club of New York intentionally lost a game to the Eckford Club of Brooklyn in 1865. This "fix" occurred four years before the first overtly professional team suited up and eleven years before the first professional league organized in North America.

Game-tampering takes a variety of forms. For example, teams and individuals have engaged in match-fixing to ensure that a given team wins. This was the case in the infamous "Black Sox" scandal in which the heavily favored Chicago White Sox threw the 1919 World Series. In recent years, Italian soccer has been repeatedly rocked by game-fixing scandals in which as many as 680 games may have been fixed (Schaerlaekens, 2013).[1] Subsequent investigations have implicated players, coaches, team owners, and referees.

Another popular form of tampering, point shaving, does not seek to alter the game's outcome. Instead, it manipulates the margin of victory. Point shaving first came to the public's attention in 1951, when a criminal investigation uncovered evidence of point shaving involving players from the University of Kentucky and several schools in New York City. Since then, periodic point shaving scandals have involved programs ranging from basketball at Arizona State University to football at Boston College. Most recently, a Detroit-area businessman pleaded guilty to bribing players at the University of Toledo to shave points between 2006 and 2008 (Dauster, 2013).

Point shaving has recently received considerable attention in both the popular press (e.g., McCarthy, 2007) and economics literature. Much of the research has been sparked by Wolfers's (2006) controversial claim that widespread point shaving could exist in intercollegiate basketball. One of the central insights of this new literature, which largely seeks to refute Wolfers's assertion, is that heavily favored teams, on average, "cover" the point spread, meaning they win by as much as bettors expect them to win. We use new techniques and data to show that Wolfers and his critics are not necessarily incompatible, that widespread point shaving could exist for regular-season NCAA college basketball games even if heavy favorites, on average, cover the point spread.

Our key contribution is that, rather than use regression techniques to compute point estimates of game outcomes, we analyze the entire distribution of outcomes. We find that, when one team is heavily favored, the distribution of game outcomes is bi-modal. Heavy favorites are likely to fail to cover the point spread or more than cover the spread. This finding is consistent with the claim that, on average, favorites win by the amount predicted by the point spread and enables us to reconcile the disputing camps in the debate. It is also consistent with Borghesi and Dare's (2009) insight that both underdogs and favorites have been found guilty of tampering with game outcomes.

This central finding is neither moral nor policy oriented. We do not claim that players or coaches are cheating on a regular basis, nor do we assert that the NCAA needs to take immediate action to clean up corruption. Instead, our contribution is statistical. We show that the commonly used statistical tools lack the power to disprove Wolfers's claim. Rather than looking at regressions and conditional means, researchers must analyze the entire distribution of game outcomes.

We also ask whether the failure to cover point spreads is the same during the regular season and during the NCAA post-season tournament play. Our results do not extend to the post-season, suggesting that team behavior is not consistent with point shaving during the post-season.

In "Literature Review," we explain point shaving and review the literature. The next section presents our empirical model and describes our data set, and in "Results," we use the distributions of game outcomes relative to point spreads to reject the null hypothesis of no point shaving for the regular season but not for tournament play. Note that rejecting this hypothesis does not prove that point shaving exists. It shows only that the outcomes are consistent with point shaving. Section five concludes.

Literature Review

Point Shaving and Basketball

Point shaving typically occurs when one or more participants in a contest, from players to coaches to referees, ensure that a team wins by less than the point spread. A point spread is the number of points by which gambling houses expect the favored team to beat the underdog. Point spreads are popular among gamblers because they attract bets on games between unevenly matched teams. A team can lose badly, but it can still pay off if it loses by less than predicted.[2]

Point shaving is more common in basketball than in low-scoring games, such as soccer or baseball. In soccer or baseball, one failure to score or one score allowed to the other team could easily determine the game's outcome. As a result, winning by less than expected could easily turn a win into a loss. Because basketball is a much higher scoring game, even close games are often decided by margins that would be considered blowouts in baseball or soccer. This, in turn, increases the opportunity to make the outcome closer than expected without altering the game's outcome.[3] Thus, we claim only that basketball is particularly susceptible to this form of corruption, not that basketball is inherently more corrupt than other sports.

Point Shaving and Gambling Markets

Economic theory treats gambling like any financial market by assuming that gambling markets are efficient. While a complete discussion of the efficient markets hypothesis is beyond the scope of this paper, efficient markets typically assume the presence of risk neutral (i.e., wealth-maximizing) individuals who make rational use of all available information. For a point spread to be efficient, the distribution of differences between game outcomes and point spreads must have a median of zero (Wolfers & Zitzewitz, 2004).[4]

Using data from the 1989-1990 through 2004-2005 NCAA basketball seasons (44,120 games), Wolfers (2006) regresses the actual winning margin on the point spread. The results, which appear in the first column of Table 2, show that the point spread is a good overall predictor of the game outcome. This finding is consistent with the hypothesis that betting markets consist of risk neutral bettors who accurately perceive the objective probability that a bet will pay off (see Vaughn Williams, 2005).

Wolfers then checks for evidence of point shaving by examining the outcomes of games with a point spread of at least 12 points. If gambling markets are efficient and the *probability density function* (PDF) of game outcomes is normally distributed, the distribution of winning margins should be symmetric and peak at the point spread. Normally distributed game outcomes—and the resulting symmetry absent point shaving—are a critical assumption in Wolfers and most subsequent research. Much of the existing literature uses tests for distributional symmetry.

To test the null hypothesis of no point shaving in college basketball, Wolfers tests whether the distribution of game outcomes is symmetric:

$$p(0 < \text{Winning margin} < \text{Spread}) = p(\text{Spread} < \text{Winning margin} < 2 * \text{Spread}) \quad (1)$$

He finds that a disproportionate percentage of heavy favorites wins the game but fails to cover the point spread (i.e., the left side of Equation (1) is significantly larger than the right side). He concludes that "point shaving led roughly 3% of heavy favorites who would have covered the spread not to cover (but still win)" (p. 282).

Borghesi (2008) presents the most significant challenge to Wolfers's interpretation. He acknowledges that too many heavily favored college teams fail to cover the point spread but notes that the same can be said of games in professional basketball and football. However, the high salaries paid to players in the National Basketball Association (NBA) and National Football League (NFL) make point shaving there highly unlikely, as players have too much to lose. (The median NBA salary was over $2 million in 2010, while the median NFL salary was close to $1 million.) Gamblers would have to offer an extremely high reward to make the risk worthwhile. As a result, the size of bets required to attract players would have to be so large that authorities would almost certainly suspect foul play. Borghesi concludes that, because the same statistical observation cannot be attributed to point shaving in professional sports, a different force must be at work in both professional and intercollegiate sports.

Borghesi asserts that heavily favored teams fail to cover point spreads because of the irrationality of bettors and the profit-maximizing response of bookmakers. He explains that the betting public "steadfastly prefers to bet on heavy favorites, even if the spread is too large" (Borghesi, 2008). Bookmakers exploit this irrationality by offering a spread that exceeds the margin they actually expect.

Borghesi's reasoning, however, applies only to favorites who do not cover the point spread. It leaves unanswered the question of why favored teams win so many games by far more than the point spread (i.e., losing teams failed to cover the spread). In fact, the hypothesis that bookmakers "shade lines" by widening the point spreads implies there should be too few blowouts, not too many. Line-shading thus resolves one problem but raises another.

Bernhardt and Heston (2010) also claim that the failure to cover has an innocent explanation. They simulate point spreads for games that have no line posted and compare the game outcome relative to this derived point spread. They find that the "distinct asymmetric patterns . . . are driven by a common desire to maximize the probability of winning" by the players (p. 15) and conclude that strategic game play causes a statistical anomaly.

Johnson (2009) points out that Wolfers's focus on heavy favorites introduces a regression effect that biases the results. This effect occurs when bettors' actions drive point spreads too high. A strong regression effect causes heavy underdogs to cover more often. Games that must have a winner and a loser (such as basketball), distort the distribution of outcomes for slight favorites: a team that is favored by two points cannot be exactly two points away from covering. This might cause researchers to reject the hypothesis of no point shaving even though point shaving does not take place.

Diemer (2009b) confirms Wolfers's findings for NFL games. He constructs and tests non-parametric PDFs of game outcomes relative to the point spreads, discarding the assumption that winning margins are distributed normally (and symmetrically). Diemer uses a less-constrained bootstrap test to determine whether the distribution of winning margins is the same for slight favorites and heavy favorites. If the gambling market is efficient, the size of the spread should not alter its accuracy as a predictor of the final outcome, so the standardized distributions of the outcomes for heavy and slight favorites should be equal. This equality, in turn, is a sufficient condition for concluding that there is no point shaving. Using this method for NFL regular-season games from 1993 to 2007, Diemer rejects the null hypothesis of equality. Instead, he finds that, "While the point spread is a function of perceived outcome, the bimodal distribution [of the heavy favorites] suggests the outcome is, in part, a function of the point spread itself. . . . As the point spread increases, so does the incentive to shave points, which leads to increasingly skewed densities [of the heavy favorites]" (pp. 22-23).

Borghesi and Dare (2009) use point spreads and the *over-under* prediction of a game's total score to calculate the expected score of favorites and underdogs in college basketball games. They then compare the predicted scores with the actual scores for games in which there is a "heavy favorite" (in the 9th and 10th deciles of the point spread) and games in which there is no clear favorite. They find that the mean scores of heavy favorites are not significantly less than expected and that heavy favorites hold their opponents to slightly fewer points than expected.

Borghesi and Dare speculate that Wolfers's findings might result from changes in strategy late in games. Coaches of heavy favorites pursue low-mean/low-risk strategies while far ahead, and coaches of heavy underdogs pursue high-mean/high-risk strategies while behind. Unfortunately, this explanation is inconsistent with rational betting markets, as bettors rapidly incorporate predictable coaching behavior into point spreads.

Bernhardt and Heston (2010) provide the richest critique of Wolfers (2006) on both the intuitive and the methodological level. Intuitively, they compare the outcomes of games that are listed by bookmakers and games that are not listed. They "find qualitatively and statistically identical patterns in the frequencies with which heavy favorites do and do not cover the spread" (p. 16). They claim that, because point shaving limits the margin of victory, informed gamblers should bet accordingly and move the point spread. Again, the outcomes of games in which the point spread moves does not differ significantly from games in which it stays relatively fixed.

On a more technical level, Berhhardt and Heston graph the distributions of outcomes for games that are—according to the above criteria—most likely and least likely to involve point shaving. They find that the distribution of outcomes does not resemble the normal distribution (though their comparison is visual and not statistical). More importantly, they find that the asymmetry in game outcomes is very similar for both sets of games, again leading them to conclude that point shaving does not cause the outcome discrepancy noted by Wolfers. This claim is reinforced by their calculations of the likelihood that heavy favorites cover the point spread. Again, they find that the probability of covering the spread is the same for games that are most likely to involve point shaving and games that are least likely to involve point shaving.

In place of point shaving, Berhardt and Heston propose two more innocent explanations for failing to cover the spread. On the one hand, "a team is far ahead may substitute backups, reducing the probability of beating a large point spread, but not a small one" (p. 23). Bernhardt and Heston note that "win maximizing strategies" dominate in a close game, as the leading team plays very conservatively and tries to use up as much time as possible, while the trailing team plays more frenetically in an attempt to score as much as possible before time runs out. Bernhardt and Heston do not say how these two contradictory strategies interact. However, the natural conclusion seems to be that raised by Borghesi and Dare (2009): Teams that are behind sometimes succeed in making the game closer (perhaps even coming back to win) by adopting a high-risk strategy. However, they frequently fail in this strategy and fall farther behind.

Bernhardt and Heston significantly advance our understanding of point shaving. However, their argument against it is not airtight. For example, because fixing games is illegal, it is successful only when undetected. Shaving points is like using performance enhancing drugs in that detection implies failure: "To succeed, the cheaters have to keep the size of the corruption small enough to prevent detection. . . . In other words, the probability of detection is endogenous" (Diemer, 2012, p. 208). It would therefore be surprising to see point spreads move for all but a very small percentage of fixed games. Similarly, if teams with big leads regularly let down at the end of games, causing the final result to be closer than it "should" be, betting markets should readily catch on and account for this in the point spread.

Two important points from our summary of Bernhardt and Heston's argument remain. Their showing that distributions do not always have the expected normal shape is important and a strong argument against point shaving. In addition, the "win-maximization" strategy by the teams involved provides an intuitively appealing alternative to point shaving.

In an interesting counterpoint to the above arguments that teams might rationally fail to cover point spreads, Paul, Weinbach, and Coate (2007) claim that favored teams have an incentive to win by more than the point spread. They assert that BCS football teams benefited in the rankings from winning by more than the point spread, particularly when the game was televised. Winning by more than the point spread could thus make the difference between playing in a BCS bowl—perhaps for the national championship—and playing in a lesser bowl. While their argument focused on intercollegiate football, the extension to basketball, where seeding for the NCAA tournament—or even making the tournament—could depend on rankings, is clear.

Paul et al. (2007) provide empirical evidence that the favored team's rankings rise when it beats the spread. Connecting their reduced form estimation to betting markets, however, requires additional structure. Again, if favored teams have an incentive to win by more than the stated point spread, then book makers should respond accordingly and adjust their point spreads to account for this incentive. Eventually, the point spread will reach a point that the favored team will not be able to exceed.

Empirical Model and Data

Data Source and Measuring Game Outcomes

We collected data from *The GoldSheet* (www.goldsheet.com) for the 1995-1996 through 2008-2009 NCAA basketball seasons. We deleted "pick 'em" games in which two teams are evenly matched (with a point spread of zero) because such games are irrelevant to our research. Because point spreads are posted only for games that are likely to attract betting, we have a sample of 31,793 regular season and 3,371 playoff games.

As noted above, we compare game outcomes to the point spread. We do so by defining *net favored points* (NFP) as

NFP = Favored team's points scored – Underdog's points scored + Point Spread (3)

where the point spread is viewed from the favorite's perspective. When the favored team covers the point spread, NFP > 0; when it fails to cover, NFP <0. If the point spread correctly predicts the outcome, NFP = 0. According to the Efficient Markets Hypothesis, the distribution of the NFP should peak where NFP = 0.

Financial Markets, Betting Markets, and the Probability Density Functions

We first conduct a parametric test of efficiency using data from the regular season by constructing kernel probability density functions (PDFs) of the NFP, using a normal optimal smoothing parameter. This builds upon the work of Wolfers (2006) and of Borghesi and Dare (2009) by comparing the actual distribution of game outcomes to the normal distribution in a statistically rigorous fashion rather than just taking an impressionistic view of the two. Figure 1 displays the PDF of all 35,164 NCAA game outcomes (both regular-season and post-season tournament games). The band around the distribution shows the 95% confidence interval around the distribution.[5] If the distribution falls within the shaded area of the reference band, the distribution is normal. As shown in Figure 1, the distribution falls within the reference band and is centered in the neighborhood of zero. This normal distribution with a peak at zero indicates that the point spread is the best aggregate predictor of the game outcomes, satisfying a necessary condition for an efficient gambling market in the aggregate.

Testing for Point Shaving in Regular-Season NCAA Games

As noted above, the incentive to shave points is greater at higher point spreads, all else equal.[6] As the spread increases, it becomes easier for interested parties to satisfy their secondary objective (winning the wager) without compromising their primary objective (winning the game). Like Diemer (2009b), we conduct a non-parametric test for equality of two distributions: heavy favorites (f(NFP)) and slight favorites (g(NFP)).[7] The formal hypothesis tests:

H0: f(NFP) = g(NFP), for all NFP

H1: f(NFP) ≠ g(NFP), for some NFP

A bootstrap test of equality generates a p-value as a global indicator of equality. We also provide a reference band around each NFP to illustrate the hypothesis test. The band is two standard errors wide at any NFP, which is measured along the horizontal axis. If the densities fall outside the reference band, that portion of the distribution is a likely reason for rejecting H0. The normal optimal smoothing parameter is the standard, as it minimizes the risk of falsely rejecting the null of equality.[8] We performed

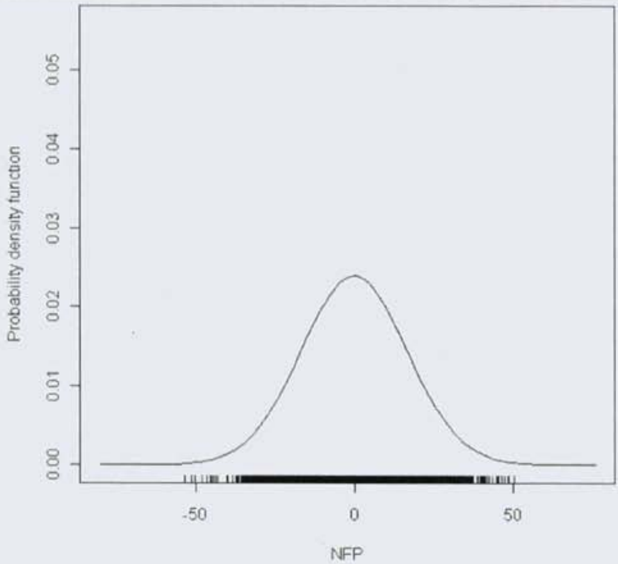

Figure 1: Empirical probability density function: Aggregate parametric data 1995-2009.

Probability distribution: Favored outcomes relative to point spread.

X axis measures the favored team's outcome relative to the point spread (NFP).

N = 35,164

Shaded area represents a normal distribution reference band.

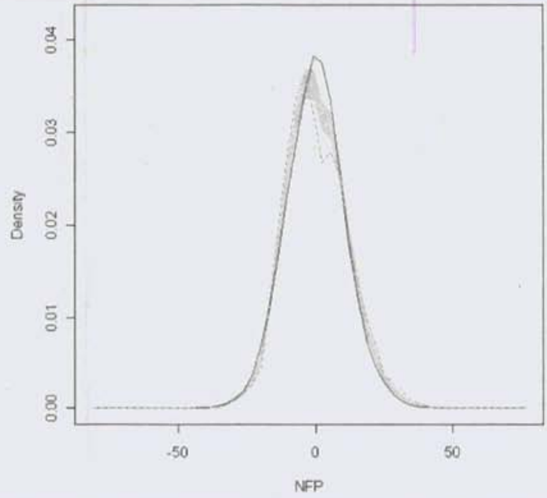

Figure 2: Regular-season test of two distributions: 19-point favorites.

Dashed line = Lower bound favorites of 19 points or more; N = 2,070.
Solid line = Upper bound favorites of 18.5 points or less; N = 29,723.
Shaded area is a reference band; the result of a bootstrap hypothesis test of equality.
Test of equal densities: p-value = 0

this test in two stages, first for 31,793 regular-season games and then for 3,371 playoff games.[9]

Results

Our results appear in Table 1 and in Figures 2, 3, 4, 6, and 7. Table 1 provides p-values of the global test for equality of the two distributions. As the first column of Table 1 shows, we reject the equality of the two distributions for every point spread in non-tournament games. The second column shows that we reject equality in tournament games only for very low point spreads but that we fail to reject equality for almost every point spread above 8.0.

Figure 2 divides regular-season games into winning margins for heavy favorites of 19 points or more (dashed line) and favorites of 18.5 points or less (solid line). The global test for equality produces a p-value of 0.00, indicating that we can reject the null hypothesis that the two distributions are the same at the one percent level. However, simply rejecting the null hypothesis that the two distributions are equal through a global test of equality (our p-value calculations) does not tell the entire story. Deeper insight would result if one could explain why the global tests failed.

The peak of the heavy favorites' PDF is at NFP = -4.3. This outcome indicates that heavy favorites win the game (they lose if NFP < -18.5) but fail to cover the spread because NFP < 0. In addition, the distribution for games with heavy favorites is bimodal, with peaks at NFP = -4.3 and 5.6 and the valley at NFP = 2.3.

The bimodal PDF for regular-season games has several important implications. First, it suggests that the point spread becomes a less precise predictor as it increases (a feature supported by our test for heteroskedasticity below). This finding is consistent

Table 1: P-values for Distributions of NFP by Point Spread

Threshold		Non-tournament games	Tournament games
Upper	Lower	p-value (N-upper, N-lower)	p-value (N-upper, N-lower)
1.0	1.5	0.00 (25,182; 1,278)	0.09 (2,645; 160)
1.5	2.0	0.00 (24,091; 2,369)	0.05 (2,499; 306)
2.0	2.5	0.00 (22,947; 3,513)	0.03 (2,347; 458)
2.5	3.0	0.00 (21,660; 4,800)	0.04 (2,183; 622)
3.0	3.5	0.00 (20,602; 5,858)	0.01 (2,037; 768)
3.5	4.0	0.00 (19,396; 7,064)	0.00 (1,880; 925)
4.0	4.5	0.00 (18,352; 8,108)	0.07 (1,753; 1,052)
4.5	5.0	0.00 (17,289; 9,171)	0.19 (1,640; 1,165)
5.0	5.5	0.00 (16,320; 10,140)	0.03 (1,513; 1,292)
5.5	6.0	0.00 (15,354; 11,106)	0.06 (1,374; 1,431)
6.0	6.5	0.00 (14,464; 11,996)	0.07 (1,266; 1,539)
6.5	7.0	0.00 (13,515; 12,945)	0.07 (1,133; 1,672)
7.0	7.5	0.00 (12,675; 13,785)	0.11 (1,040; 1,765)
7.5	8.0	0.00 (11,777; 14,683)	0.09 (931; 1,874)
8.0	8.5	0.00 (11,008; 15,452)	0.04 (842; 1,963)
8.5	9.0	0.00 (10,223; 16,273)	0.48 (726; 2,079)
9.0	9.5	0.00 (9,514; 16,946)	0,31 (658; 2,147)
9.5	10.0	0.00 (8,894; 17,566)	0.34 (589; 2,216)
10.0	10.5	0.00 (8,213; 18,247)	0.21 (537; 2,268)
10.5	11.0	0.00 (7,624; 18,836)	0.46 (482; 2,323)
11.0	11.5	0.00 (7,034; 19,426)	0.58 (440; 2,365)
11.5	12.0	0.00 (6,500; 19,960)	0.39 (385; 2,420)
12.0	12.5	0.00 (6009; 20,451)	0.11 (344; 2,461)
12.5	13.0	0.00 (5,520; 20,940)	0.03 (312; 2,493)
13.0	13.5	0.00 (5,038; 21,422)	0.09 (277; 2,528)
13.5	14.0	0.00 (4,599; 21,861)	0.32 (247; 2,558)
14.0	14.5	0.00 (4,154; 22,306)	0.58 (218; 2,587)
14.5	15.0	0.00 (3,769; 22,691)	0.64 (196; 2,609)
15.0	15.5	0.00 (3,416; 23,044)	0.30 (172; 2,633)
15.5	16.0	0.00 (3,107; 23,353)	0.58 (157; 2,648)
16.0	16.5	0.00 (2,796; 23,664)	0.49 (141; 2,664)
16.5	17.0	0.00 (2,561; 23,899)	0.42 (128; 2,677)
17.0	17.5	0.00 (2,300; 24,160)	0.37 (122; 2,683)
17.5	18.0	0.00 (2,065; 24,395)	0.29 (114; 2,691)
18.0	18.5	0.00 (1,854; 24,606)	0.42 (101; 2,704)
18.5	19.0	0.00 (1,687; 24,773)	0.33 (97; 2,708)
19.0	19.5	0.00 (1,512; 24,948)	0.38 (84; 2,721)
19.5	20.0	0.00 (1,384; 25,076)	0.58 (80; 2,725)

with Diemer's (2009b) finding a bimodal distribution of outcomes relative to the spread in the NFL regular-season games. While means are valuable summary statistics, they cannot tell us about the entire distribution of scores. In particular, a mean of zero—which is frequently taken as indicating no point shaving—is consistent with a bimodal distribution that has peaks on either side of zero. Such a distribution suggests that big favorites either cover the point spread by too much or fail to cover at all.

The asymmetry of the PDF also counters the claim that rejecting the null hypothesis of identical distributions is a statistical anomaly that results from strategic effort (Bernhardt & Heston, 2010), line shading (Borghesi, 2008), or regression to the mean (Johnson, 2009). If the results are a statistical anomaly, the PDF would still be symmetric. However, as shown in Figure 2, the heavy favorites' PDF (the dashed line) is far from symmetric. In addition, Figure 2 shows that the PDFs fall outside the reference band when 24.3>NFP>15.8. This supports the finding that "heavily favored teams appear to be involved in too many blowouts" (Wolfers, 2006, p. 282). In other words, heavy favorites either win by far more than the point spread or fail to cover (but still win).

This finding appears to contradict a basic observation: most point shaving scandals involve close games that frequently result in losses for the favored team.[10] This observation need not contradict our results, as scandals necessarily involve point shaving that has been uncovered. Our estimation takes a more global view and shows that there is more evidence of behavior that is consistent with point shaving at large point spreads. A rational gambler would want to shave points in these games precisely because the chance of discovery might be greater when shaving points affects the game's outcome.

The distortion of the distribution to the point that it becomes bimodal shows that teams failing to cover the point spread is not the result of bookmakers' mistakenly setting the point spread too high (due to such factors as line shading or strategic game play). Figure 2 shows that, while an unusually large percentage of heavy favorites fails to cover the point spread (the left peak), an unusually large percentage of heavy favorites also wins by unexpectedly large margins (the right peak).

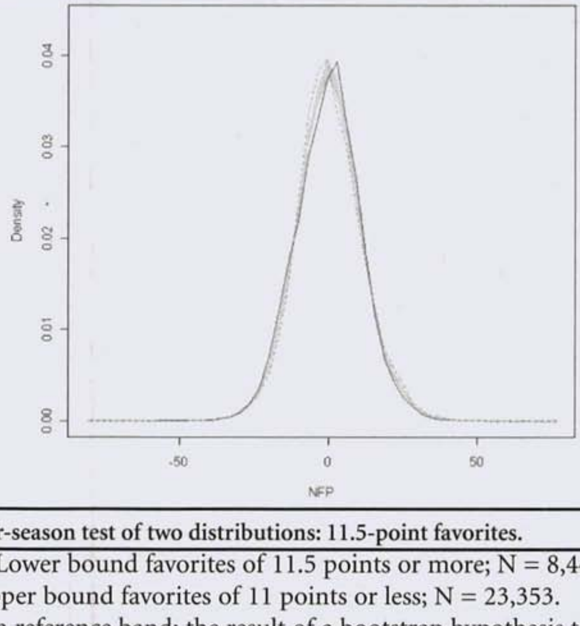

Figure 3: Regular-season test of two distributions: 11.5-point favorites.

Dashed line = Lower bound favorites of 11.5 points or more; N = 8,440.
Solid line = Upper bound favorites of 11 points or less; N = 23,353.
Shaded area is a reference band; the result of a bootstrap hypothesis test of equality.
Test of equal densities: p-value = 0

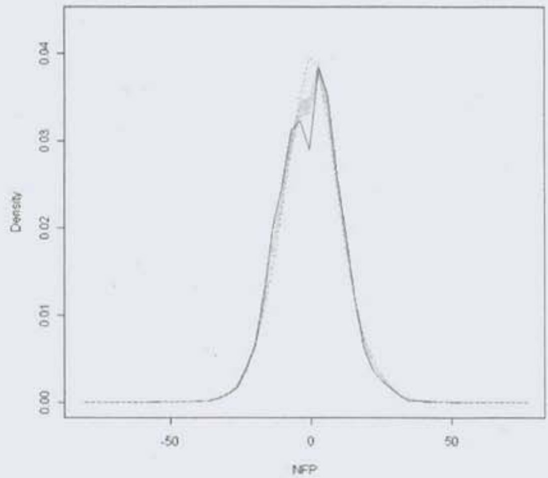

Figure 4: Regular-season test of two distributions: 3.5-point favorites.

Dashed line = Lower bound favorites of 3.5 points or more; N = 24,730.
Solid line = Upper bound favorites of 4 points or less; N = 7,063.
Shaded area is a reference band; the result of a bootstrap hypothesis test of equality.
Test of equal densities: p-value = 0

Figure 3 displays the results at the 11.5-point threshold. The p-value is again 0.00, meaning we again reject H0. This time, each PDF peak is located outside the reference band at NFP= -1 (for the heavy favorites) and NFP = 2.5 for the slighter favorites. The heavy favorites are again involved in too many blowouts at 27.8>NFP >19. Finally, the PDFs (especially the heavy favorites, as denoted by the dashed lines) become increasingly skewed as the point spread increases. Behavior consistent with point shaving thus increases as the point spread increases.

It is also possible for the distribution of outcomes to be distorted by the no-tie-game constraint. Because basketball games cannot end in a tie, the PDF for games with relatively slight favorites is skewed. Figure 4, which shows teams favored by 3.5 or more points and teams favored by 3 points or less, highlights this fact, resulting in a p-value of 0. For example, games with 3-point favorites cannot have outcomes for which NFP = -3. The distribution of outcomes for relatively small favorites (the solid line) is bimodal due this constraint. The peaks are NFP = -3.9 and 2.5, and the valley is at NFP = -0.7, resulting in negative kurtosis for weaker favorites. If we could correct for the no-tie-game constraint in Figure 2, the density function of the slighter favorites would have a single, higher peak. In other words, the games that create the solid, bimodal distribution shown in Figure 4 are also in the sample that produces the solid distributions in Figures 2 and 3, thus flattening those distributions. The no-tie-game constraint thus causes us to understate the evidence of point shaving.

Heteroskedasticity and Point Shaving in Regular-Season Games

Our results thus far reject the hypothesis that the PDFs of the NFP for heavy favorites and weak favorites are identical in regular-season games. However, the bootstrap tests of equality are highly sensitive to sample size changes. To avoid inappropriately rejecting (or failing to reject) the null hypothesis of no point shaving, we run tests that are less sensitive to sample size problems. We do so by reproducing Wolfers's basic equation for our sample of regular-season NCAA games. As noted above, Wolfers finds that the actual point differential rises roughly 1:1 with the point spread. He does not, however, examine the variance of this regression. A constant variance would be consistent with the null hypothesis of no point shaving, but a variance that changes with the point spread would be further evidence that the point spread affects the margin of victory.

We perform this test by regressing the ex-post point differential on the point spread for all 31,793 regular-season games in our sample. We then test for heteroskedasticity. If we cannot reject the null hypothesis of homoskedasticity, then we have an additional argument that point shaving does not exist. Failing to reject the null hypothesis provides further evidence of behavior that is consistent with point shaving.

The results of the regression appear in the second column of Table 2. While our results are not identical to Wolfers's, we find that the point differential rises roughly 1:1 with the point spread. As before, this result is consistent with risk-neutral individuals who correctly perceive game outcomes in the aggregate. The key finding in the first column of Table 2, however, is not in the coefficients or R^2. Instead, it is in the result of the Breusch-Pagan test for heteroskedasticity. The test rejects the null hypothesis of homoskedasticity for regular-season games at the one percent level of significance. This result indicates that the distribution of the error structure varies with the point spread for regular-season games, which supports rejecting the null hypothesis of no point shaving.

Tournament Play

We next perform the same tests using game outcomes in post-season tournament play. We test separately because the consequences of losing in tournament play are much more serious than losing during the regular season, as a loss generally results in the end of the team's season and the end of many players' collegiate careers. The pressure to win a tournament game is therefore much greater than for almost any regular-season

Table 2: Regression of Point Differential on Point Spread

Coefficient	Wolfers	Regular season	Tournament
Point spread	1.007	1.048***	1.045***
	(167.83)	(104.23)	(29.92)
Constant	-0.012	-0.406***	0.503*
	(0.20)	(3.94)	(1.70)
Adjusted R-squared	0.39	0.26	0.21
Breusch-pagan Chi-squared	N/A	8.17***	0.83

t-statistics in parentheses
*Significant at 10% level
**Significant at 5% level
***Significant at 1% level

game. This should, according to the reasoning of Borghesi and Dare (2009) and Berhardt and Heston (2010), increase the intensity of play at the end of tournament games, except perhaps for those where the outcome is out of hand. Teams that are behind can therefore expect their high-risk strategies to result more extreme outcomes, games that are even closer or even less competitive. This reasoning thus predicts that the twin peaks of the bimodal distribution should become even more pronounced for tournament games.

Games that are not close at the end could result in less effort by both teams. The losing team might be disheartened, while the coach of the winning team rewards lesser players by giving them extended playing time.[11] If both teams reduce effort, the impact on the game outcome is uncertain. However, if reduced effort systematically swings the outcome one way or the other, this result should be reflected in the point spreads.

Our test shows whether our findings for regular-season games extend to post-season tournaments. If there is no point shaving, all available information about the game outcome is incorporated in the point spread for regular-season and post-season games, and the PDFs of the NFP should be the same for both. Even if point shaving occurs during the regular season, we expect it to be less likely to occur in tournament games. Because tournament games are more important than regular-season games, the cost of losing a game is greater, and players and coaches will be more averse to behavior that increases the risk of losing. In addition, the greater attention that the public pays to tournament games increases the probability that point shaving will be detected.

Our sample consists of 3,371 tournament games. Tournaments are classified as any potentially season-ending games, such as the NCAA Tournament, NIT Tournament,

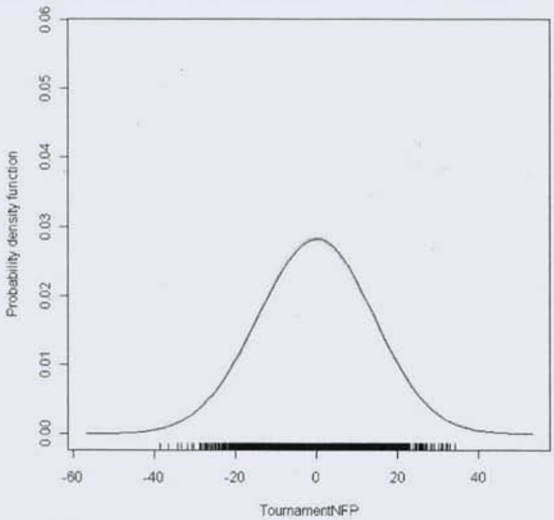

Figure 5: Probability distribution: Favored outcomes relative to point spread tournament play.
X axis measures the favored team's outcome relative to the point spread (NFP).
N = 3,371.
Shaded area represents a normal distribution reference band.

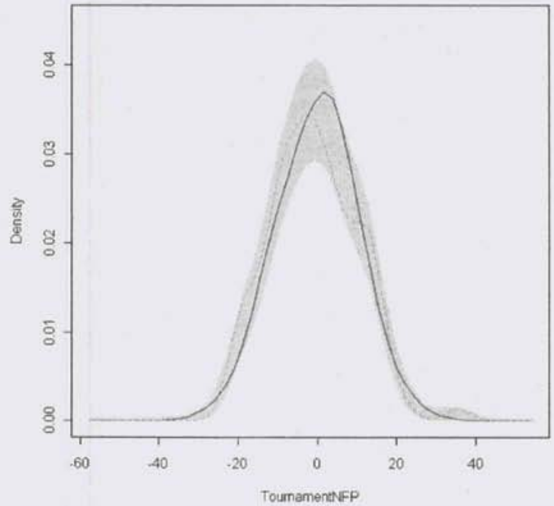

Figure 6: Tournament play test of two distributions: 19-point favorites.
Dashed line = Lower Bound favorites of 19 points or more; N = 99
Solid line = Upper bound favorites of 18.5 points or less; N = 3,272
Shaded area is a reference band; the result of a bootstrap hypothesis test of equality.
Test of equal densities: p-value = 0.28

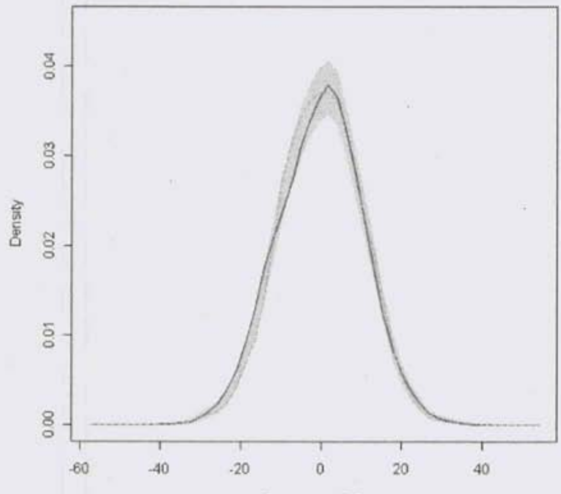

Figure 7: Tournament play test of two distributions: 11.5-point favorites.
Dashed line = Lower bound favorites of 11.5 points or more; N = 520.
Solid line = Upper bound favorites of 11 points or less; N = 2,851.
Shaded area is a reference band; the result of a bootstrap hypothesis test of equality.
Test of equal densities: p-value = 0.35.

and Conference Tournaments. We again construct and test the PDFs and perform the global test of equality for the PDFs of strong and weak favorites.

Results appear in Figures 5, 6, and 7. Figure 5 shows that the aggregate distribution is normally distributed, as the distribution falls inside the confidence band. Figures 6 and 7 test whether the PDFs of heavy and slight favorites are equal. As before, the point spread thresholds are 19 and 11.5 points. The resulting p-values –0.28 and 0.35 – do not allow us to reject H0. The figures show that the distributions fall inside the reference band, so any fluctuations probably reflect white noise.

Heteroskedasticity and Point Shaving in Tournament Games

We next ran Wolfers's basic OLS equation for NCAA tournament games. Running this test separately for tournament games also indicates whether teams behave differently in post-season tournaments. If there is no change in behavior, then these results should resemble those for regular-season games. As noted earlier, the bootstrap tests' sensitivity to sample size may lead to false negatives, failing to reject the null of equality when differences do exist. To ensure that these results are not the result of sample size fluctuations, we test for heteroskedasticity. The heteroskedasticity test incorporates all 3,371 tournament games, thus eliminating sample size concerns.

If behavior is the same in the regular season and post season, we should find heteroskedasticity in our estimates of post-season games, with weaker predictive power at larger point spreads. If we cannot reject the hypothesis of a constant variance, then we have reason to believe that the incentives in the post-season differ from those in the regular season. This time, as the incentives to shave points decreases in tournament play, statistical evidence of point shaving (heteroskedasticity) disappears.

The results of this regression appear in the third column of Table 2. As before, we find that the actual point differential rises 1:1 with the point spread. This time, however, the Breusch-Pagan fails to reject the null hypothesis of constant variance. By eliminating sample size concerns, this finding provides further evidence that the distributions of the NFP is constant regardless of the point spread and that point shaving does not occur for tournament games.

Conclusion

Wolfers's (2006) claim that point shaving exists in college basketball has drawn considerable attention in the literature, most of it attacking his findings. The result has been a sizable literature in response to Wolfers's finding. Like much of this literature and like conventional wisdom, we find no evidence of rampant point shaving in intercollegiate basketball. However, we do show that many of the seemingly conflicting results can be nested in a broader approach that accommodates many of the existing findings.

Our key finding is that the distribution of game outcomes is bimodal. A bimodal distribution with one peak on one side of the "no corruption" outcome and one peak on the other side is consistent with point shaving by both favorites, who lose by too little, and underdogs, who lose by too much. At the same time, the mean outcome is consistent with a world in which there is no shaving and teams follow win-maximizing strategies.

If teams attempt to maximize wins, then much of the previous literature suggests that the bimodal distribution should become even more pronounced for tournament games, in which the pressure to win is much greater than for regular-season games. However, we find that just the opposite occurs, and the distribution of game outcomes is indistinguishable from a normal distribution.

The finding that the distribution of game outcomes reverts to a normal distribution suggests that there is no widespread point shaving in tournament games. This result may reflect the fact that rational agents respond in the expected manner to incentives, as the greater attention paid to tournament games makes detection more likely than for regular-season games.

We show that there is greater evidence of point shaving for regular-season games that have large point spreads. This satisfies gamblers who want to fix the game while minimizing the chance that the favored team will lose the game. Because post-season games receive much closer scrutiny than regular-season games, they are likely to be much harder to fix undetected. Moreover, because post-season games are much more important than typical regular-season games, we expect all the parties involved to be more reluctant to participate in point shaving. Our results confirm this hypothesis. We thus reject the null hypothesis of no point shaving when the incentives to shave points are strong and fail to reject the null hypothesis when the incentives are weak.

References

Bowman, A., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis: The kernel approach with S-plus illustrations.* London: Oxford University Press.

Bernhardt, D. & Heston, S. (2010). Point shaving in college basketball: A cautionary tale for forensic economics. *Economic Inquiry, 48,* January, 14-25.

Borghesi, R. (2008). Widespread corruption in sports gambling: Fact or fiction? *Southern Economic Journal, 74,* 1063-1069.

Borghesi, R. & Dare W. (2009). A test of the widespread point-shaving theory. *Finance Research Letters, 6,* September, 115-121.

Brown, R. (1994). Measuring cartel rents in the college basketball recruitment market. *Applied Economics, 26,* January, 27-34.

Dauster, R. (2013). Businessman pleads guilty in Toledo point-shaving scandal. NBC Sports: College Basketball Talk. Retrieved from http://collegebasketballtalk.nbcsports.com/2013/02/05/businessman-pleads-guilty-in-toledo-point-shaving-scandal/

Diemer, G. (2009a). Internet sports books operating in an inefficient marketplace. *The Journal of Gambling Business and Economics, 3,* 1-22.

Diemer, G. (2009b). Point shaving in the NFL gambling market: A bootstrap investigation. *Journal of Prediction Markets, 3,* December, 13-31.

Diemer, G. (2012). Comment on "Investigating allegations of point-shaving in NCAA basketball using actual Sportsbook betting percentages." *Journal of Sports Economics, 13,* 207-210.

Johnson, N. (2009). NCAA point shaving as an artifact of the regression effect and the lack of tie games. *Journal of Sports Economics, 10,* February, 59-67.

Leeds, M. A. & von Allmen, P. (2010). *The economics of sports.* Boston, MA: Pearson.

McCarthy, M. (2007). Point shaving remains a concern in college athletics. *USA Today.* Retrieved from http://www.usatoday.com/sports/college/2007-05-08-point-shaving-cover_N.htm

Paul, R. & Weinbach A. (2011). Investigating alligations of pointshaving in NCAA basketball using actual Sportsbook betting percentages. *Journal of Sports Economics, 12,* August, 432-447.

Paul, R. & Weinbach, A. (2012). Investigating alligations of pointshaving in NCAA basketball using actual Sportsbook betting percentages: Response. *Journal of Sports Economics. 13,* April, 211-217.

Paul, R., Weinbach, A., & Coate, P. (2007). Expectations and voting in the NCAA football polls: The wisdom of point spread markets. *Journal of Sports Economics, 8*, August, 412-424.

Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature, 36*, December, 2021-2064.

Schaerlaeckens, L. (2013). Soccer's credibility crisis deepens. Fox Soccer. Retrieved from http://msn.foxsports.com/foxsoccer/world/story/soccer-credibility-crisis-deepens-as-match-fixing-scandal-continues-021313

Seymour, H. (1960). *Baseball: The early years.* London: Oxford University Press.

Vaughan Williams, L. (Ed.). (2005). *Information efficiency in financial and betting markets.* Cambridge, England: Cambridge University Press.

Wolfers, J. (2006). Point shaving: Corruption in NCAA basketball. *American Economic Review, 96*, 279-283.

Wolfers, J. & Zitzewitz, E. (2004). Prediction markets, *Journal of Economic Perspectives, 18*, 107-26.

Endnotes

[1] The most recent scandal uncovered by the European police intelligence agency suggests 680 games were suspicious within its 19-month investigation.

[2] In a variant of point shaving, some gamblers have tried to affect the total number of points scored in a contest, a figure known as the over-under.

[3] Most studies assume that point shaving is a function of the effort level of the favored team. We return to this point later.

[4] The degree of efficiency depends in part on whether some bettors have access to more information than others. For more information on point spreads see Diemer (2009A). For more information on efficiency in gambling markets, see Sauer (1998) and Vaughn Williams (2005).

[5] The shaded area is barely visible in this instance.

[6] Paul and Weinbach (2011, 2012) assert that all things are not equal. They point to the correlation between heavy favorites and a disincentive to shave since the heavy favorites would have the players who are likely to jeopardize professional careers by point-shaving.

[7] Eliminating the parametric assumption of the dataset eliminates an added constraint. Instead of testing for parametric characteristics of subsets of a dataset we test the equality of two subsets.

[8] Using a Sheather-Jones plug-in bandwidth yielded similar results.

[9] Page constraints limit the findings presented here. All point spread thresholds are available upon request.

[10] We are grateful to an anonymous referee for pointing this out.

[11] We thank an anonymous referee for pointing this out.

Authors' Note

We thank James Bailey, Michael Bantner, Erwin Blackstone, Brad Humphreys, Randy McClure, and the participants in the Temple University Economics Seminar for their helpful comments and suggestions.