

**Point Shaving in the NBA:
An Economic Analysis of the National Basketball Association's Point Spread
Betting Market**

May 11th 2007

Jonathan Gibbs
jmgibbs@stanford.edu
Department of Economics
Stanford University, Stanford, CA 94309

Advisor: Roger Noll

Abstract

In this paper, I examine the point spread betting market of the National Basketball Association for indications of cheating through point shaving. Previous studies have illustrated the contradictory phenomenon of simple, profitable betting rules persisting in a market of informed traders. While normally dismissed due to the salaries earned by professional players, I propose point shaving as an explanation for this persistent market inefficiency. Through analysis of the distribution and symmetry of the forecast errors, I arrive at a rough approximation of point shaving's prevalence. In studying in-game scores, I find conclusive evidence that the point spread influences game outcomes and has causative effects on the probability of the favored team covering the point spread. The key economic concept these results support is that with the non-linear payout structure of the point spread betting market creating a discontinuity over the value of final game margins, individuals are altering their behavior in response to these distorted incentives.

Keywords: sports betting, gambling market, NBA, sports book, efficient markets, wager

I am indebted to Roger Noll and Nick Bloom for their insights and patience throughout this process. I would like to thank my draw mates for helping to shape and listening to my ideas, and my parents, grandparents, sister and girlfriend for their help, support, love, and encouragement. All errors are my own.

I. Introduction

Sports gambling markets have long been used by economists to study and test hypotheses about the operation of financial markets. Among the useful attributes of betting markets is that they have a definitive endpoint at each game's conclusion, yielding an objective value of the traded asset. Point spread betting markets also provide predictions, in which prices represent forecasts of the final point differential of a game. Data on betting lines and game outcomes have been used in recent empirical financial research to study questions such as how markets incorporate information and the factors affecting price variations. A literature review indicates that the overall performance of the National Basketball Association (NBA) betting market is efficient, with public and private information incorporated into prices. This paper focuses on a segment of the NBA betting market that may be susceptible to the equivalent of "insider trading." For basketball, this phenomenon manifests itself as "point shaving." In return for a bet's monetary payoff, a point shaving player chooses to affect his play negatively in order for his team to still win, but by a smaller margin than the predicted point spread.

When discussing the possibility of point shaving in the NBA, the overwhelming majority of basketball fans immediately point to the large contracts players are paid as evidence against this form of cheating. In ESPN.com writer Chris Sheridan's blog post of February 14, 2007, he recalls asking NBA commissioner David Stern, "if the potential for corruption -- players fixing games or shaving points -- was a significant concern of his. [Stern's] answer was no." In their analysis of the NBA betting market, Dobra, Cargill, and Meyer (1990) succinctly stated that "although [they] recognize that 'fixes' are not impossible to arrange, . . . they are very unlikely, and therefore, the [betting] lines . . . are probably not substantially affected by this kind of activity" (p.247). It is true that the NBA has become a billion dollar global business and that for

the 2006–2007 season, players averaged an annual salary of \$5.215 million, a 1,475%¹ increase since the 1984–1985 season (Ford 2006). However, simply because an individual already earns a high wage, does not imply that the potential to earn additional money will not be attractive. There is a disincentive associated with the potential for lost wages if a player is caught shaving points, however two other factors influencing player salaries are not often mentioned.

First, the demand for NBA contracts far exceeds their supply. The total number of NBA roster spots is fixed at a maximum of only 450, (thirty teams each with fifteen players). The combination of a large player pool and few available contracts engenders substantial market competition, with past NBA experience not close to a guarantee of future earnings. Second, even the most competitive and talented players are not assured of long term NBA success². Basketball is a physically and mentally demanding sport, with injuries almost inevitable. Even small medical issues can bring a player one step closer to retirement, implying these players' world class skill only has earnings potential for a limited duration. Aside from the risk of injuries, this future earning potential is further negatively impacted by expected declines to a player's skills resulting from aging. The sharp decrease in expected wages accompanying players' retirements emphasizes the limited window of opportunity. The uncertainty created by these two factors works to somewhat counteract the decreasing marginal utility of money argument described above.

There is also anecdotal evidence suggesting players already influence their play for financial gain. NBA players constantly face accusations of playing to accumulate *individual* statistics instead of playing to elevate the *team's* chances of winning. This "me-first" attitude designed to capitalize on the dollars available on the free agent market further separates the player's goals from that of the team. The "contract-year-phenomenon" provides similar

anecdotal evidence suggesting that players in the final year of their contract outperform their past statistical averages in order to earn a larger contract at year's end. Once signed to a new contract, players then revert to their pre contract-year level of production. This boost in performance and subsequent regression demonstrates the "me-first" attitude, as teams should expect players to always perform at their peak, and demonstrates players altering their level of play in response to monetary incentives.

Similarly responding to large monetary incentives, in 2007, the NBA held its annual all-star game in Las Vegas for the first time. While one major stipulation for having the game in Las Vegas was that all sports books could not offer gambling action on the game, NBA players and gambling have gone hand in hand for a long time³. ESPN.com columnist Bill Simmons in his June 19, 2006 article for ESPN The Magazine goes so far as to claim that a predisposition to gambling is a positive trait within a basketball player. The intense desire to compete and more importantly the overwhelming need to win characterize the best basketball players, separating them from the rest of the league. He posits that the fortitude to put everything on the line and take risks in an effort to win is as much a necessity in basketball as it is in high-stakes gambling. Michael Jordan, arguably the best basketball player ever, was a notorious gambler during his career. There are a multitude of stories with regards to Jordan and his gambling, most often concluding, similarly to Simmons, that his gambling demonstrates his legendary drive to compete and all-consuming need to win. The similarity in requisite instincts and skills between the two activities makes the transition from basketball to gambling seamless.

The question thus becomes, have players crossed the line, and begun betting on NBA games? Sports gambling has enjoyed a renaissance in the past fifteen years with the internet creating an explosion of betting sites offering daily NBA betting lines, or point spreads. No

longer restricted to the sports books of Las Vegas, annual betting volumes in the millions are placed online and around the world. Bookies attempting to pay off players to fix games are obsolete as players can with far more ease simply place a bet themselves. This creates a dangerous intersection as the game making these players rich is also a bettor's haven. On March 20, 2007, in a 100–98 road loss against the Portland Trailblazers, all–star guard Gilbert Arenas of the Washington Wizards was involved in a betting incident. He made separate bets with two different fans during the game for \$10, in which if he had the opportunity to take a game winning shot, he would win the bet if he made that final shot; Arenas got his chance, missed, and paid up (Associated Press 2007). He was quickly admonished by the NBA when the bets were made public after the game, but it furthers speaks to the connection between basketball players and gambling. These players are young, hyper–competitive, rich, attracted to gambling and armed with better information over NBA game outcomes than the odds–makers themselves. As these players have superior knowledge of the sport and more importantly, a sphere of influence over game outcomes, simply setting a betting line provides NBA players with incentives to cheat.

II. The NBA Betting Market

The NBA betting market needs to be addressed before asking the question of how to detect whether NBA players are cheating for monetary gain. For any NBA game a variety of factors affect each team's chances of winning, including the teams' relative strengths, which team is at home, the results of each team's previous games, luck, as well as a host of other elements. Based on an analysis of these factors, one team usually has a greater chance of winning. Thus, in order for individuals to bet on one team or the other, odds–makers have two options for creating a betting market. They can either decrease the payout for those betting on the team more likely to win, or they can handicap each team's chances of winning by

announcing a betting line. Both options exist within the betting market with the former being less relevant as the latter is much more prevalent. This analysis limits its focus to solely the betting line market as the discontinuity in its payout structure discussed in the next section provides the incentive for players to cheat.

As an example, for the NBA game on March 18, 2007 between the Detroit Pistons and the Dallas Mavericks, the betting line was set at the Mavericks being favored by 5.5 points. In this instance, a bet on the favorite, the Mavericks, wins if the Mavericks win by more than 5.5 points; conversely, a bet on the underdog, the Pistons, wins if the Pistons win OR if the Mavericks win by less than 5.5 points. Since NBA scores increase by integers, while betting lines are set in 0.5 point increments, this betting line was guaranteed to produce winning and losing bets. If instead the Mavericks had been favored by 6 points and the outcome of the game was such that the Mavericks won by exactly 6 points, the bet would have “pushed,” and bets for both sides would have been returned in full.

The question of how the point spread is set is interesting. The Las Vegas Sports Consultants (LVSC), a company owned by Michael “Roxy” Roxborough establishes the odds for approximately 75% of the licensed sports books in Nevada, with most illegal and online books in and out of Nevada drawing their odds from what the LVSC develops (Stern 1999). These betting lines are posted at 8:00 AM Pacific Standard Time (PST) on game day and fluctuate from market trading until each game’s tip-off, at most eleven hours later. The company’s stated goal and the generally accepted assumption is that these opening betting lines are set to equalize the amount of money bet on each side of the betting line, minimizing the bookmakers’ financial risk and guaranteeing their revenue. (Whether this assumption holds is discussed in Section IV, the literature review.) This riskless profit is assured by the “eleven-for-ten” trading rule where a

standard fee (the vigorish) is taken from all winnings; the price the bookmakers charge for creating the betting market. By taking in 100% of the money from losing bets while only paying out 90.91% of the money from winning bets, as long as there is an equal amount of money bet on either side of the betting line, the bookmaker profits regardless of the game's outcome. By attempting to balance the dollar amount bet on each team, if the NBA betting market is efficient, the point spread corresponds to the number of points by which the favored team is expected to win.

If the opening betting line does not attract equal monies bet on either side of the betting line, the sports book often moves the point spread during the eleven hour period between opening and closing to avoid risk. Unlike in pari-mutuel betting where odds are recalculated as bets are placed, point spread bets lock in the betting line at the time of the bet's placement. Thus if the Mavericks were underpriced at their opening betting line (only favored by 3 points, when the optimal point spread is 5.5 points), there will be excess demand for the Mavericks and insufficient demand for the Pistons. This deviation from the optimal betting line signals a profit opportunity to the betting public at the expense of the bookmaker, and would induce betting line changes in response to market activity. The closing betting lines therefore reflect the dollar-weighted average of the public's opinion on the outcome of the game, not the bookmakers. The point spread can thus be viewed as a market clearing price and a forecast of a game's final score, with the difference between the actual outcome and the betting line's prediction interpreted as a forecast error. If NBA players are cheating, negatively affecting their play to win bets, these forecast errors could demonstrate an observable bias relative to the point spread.

III. Point Shaving Conceptually

There are three methods through which players could influence their play in order win outside wagers. The first and most prominent method is by throwing games. In baseball, eight players on the Chicago White Sox of 1919 took bribes during the World Series in return for purposefully losing games. Similarly, NBA players could negatively alter their play to this level inducing their teams to lose strategic games on which they have placed bets. However, the poor level of play necessary to pull off this feat could have problematic ramifications for the player. First, the player could be benched by the coach due to his poorer performance, preventing his influence on the game's outcome. Second, his poor play could impact future contracts and thus future wages. Third, part of the reason players are attracted to gambling and placing bets is their hyper-competitive nature, which would go unfulfilled resulting from their team's loss. These three aspects make the possibility of throwing games unappealing.

The second method by which players could alter their level of play for monetary gain is by exceeding the player's expected effort level. If the betting public expects a player to play at a certain level of intensity and sets the betting line accordingly, the player could choose to exceed that effort level, having placed a bet on his own team to win and cover the point spread. The extra unexpected effort the player exerts would be unaccounted for in the betting market and would make his bet's outcome more likely. This theory is possible and could be demonstrated in the data; however it does not seem probable. First, players already have strong monetary incentives to exert as much effort as possible in order to earn as large a contract as possible. Second, after increasing one's effort for a specific game, there may be negative effects from the coach or fans when in subsequent games the player's effort returns to previous levels. Third, it is

much less appealing to increase one's effort, when a slight decrease in effort to instead shave points would have the same monetary result.

The third method, the practice of point shaving, exploits the discontinuity in the betting line's payout structure. While the betting line is designed to balance the betting market, it does not correspond to the incentives of the teams: the teams only care about wins and losses. This asymmetry creates a window of space in which the favored team can still win the game while a bet on the favored team loses. This window is an opportunity for players to shave points without adversely affecting their teams' wins and losses. If done carefully, a player on a favored team could slightly reduce his effort to change the expected final margin without changing his team's likelihood of winning. This decrease in effort, done in conjunction with the placement of a bet on the team's opponent to cover the point spread, is point shaving. In this scenario, the player gets the best of both worlds; his team wins the game and he receives a financial payoff as a result of purposely slightly underperforming.

As shown in Figure 1, if the player is exerting expected maximum effort, the probability of just failing to cover the point spread (B_1) equals the probability of just covering the point spread (C_1). However, by slightly reducing his effort he can induce the game to end up in (B_1) instead of (C_1) without changing the outcome of the game, solely shaving points off the final margin of victory. In essence, the betting market assigns values to final margins with a discontinuous jump in value occurring at the point spread. As the final margin has little if not no intrinsic value, (other than determining whether the team wins or loses), the betting market's discontinuous valuation of final game outcomes provides incentives to arrive on one side of the betting line relative to the other. Thus, as one's effort is more easily decreased than increased,

point shaving is the response to these monetary incentives fundamentally created from the setting of the betting line.

Once again, the typical response to the possibility of point shaving is that the large salaries NBA players earn should counteract the incentive to cheat for monetary gain. However, as was discussed above and through further breaking down the decision making process in equations (1) and (2), this argument does not seem satisfactory.

$$\text{Benefits: (Point Shaving Success Rate)*(Betting Payoff)} \quad (1)$$

$$\text{Costs: (Lost Future Income + Lost Reputation)*(Probability of Detection)} \quad (2)$$

In comparing the two equations, since the (Probability of Detection) is near zero, the expected costs of point shaving are tiny relative to the expected benefits: $(1) > (2)$. Additionally, by selecting only certain games in which to engage in point shaving the player is able to maximize the (Point Shaving Success Rate), while further minimizing the risk of detection. Games in which the player's team is heavily favored maximize the window in which the team can win, while a bet on the underdog to cover pays off, or in returning to Figure 1, maximizes the area in region B. Thus, while each team's overriding goal is to win each game, the incentive to shave points is greatest for games in which one team is heavily favored. If point shaving is occurring, this theory suggests that it is in these games the practice would most likely be observed.

At the college basketball level, where players are not paid, there have been three major point shaving, cheating, scandals in the past twenty-five years, two occurring in the past ten years. Recent statistical analyses of past college game results indicate that this point shaving problem may be more serious and widespread than is generally believed (Wolfers 2006). Yet, despite college players making the annual transition to the NBA, there has never been a point shaving scandal in the history of the NBA. Conventional wisdom ascribes the hefty salaries

NBA players are paid as sufficient to eliminate any temptation to shave points for financial gain. However, the NBA's expensive culture, the players' limited and uncertain earning potential, and the players' ultra-competitive nature, combined with the miniscule likelihood of being caught may counterbalance this mainstream thinking. Furthermore, even if personal (or friend/familial) financial motivations are absent, the overpowering need to win or need to satisfy one's ego, may still make betting on NBA games attractive.

The focus of this analysis is to detect whether NBA players have in fact crossed the line and begun betting on their own games. Utilizing data over betting lines, in-game scores, and game outcomes, I propose to investigate the suspicion that gambling has caused NBA players to affect their play to influence the outcome of NBA games in order to win bets. If this phenomenon is occurring, it should leave observable anomalistic footprints in the data. Thus, the question asked by this empirical study is whether there are statistical indications of past point shaving in the NBA. This paper proceeds with a discussion of the relevant literature and their examinations of the NBA betting market including a discussion of how this analysis shall enhance the existing literature and attempt to answer the described question above. The next section details the econometric and statistical methods used as well as the data. An in-depth discussion of the results follows before concluding remarks.

IV. Literature Review

The majority of the scholarly literature discussing this subject focuses on (1) revealing the subtle inefficiencies in a mostly efficient NBA point spread betting market and then (2) attempting to provide an explanation as to why these inefficiencies persist. It has become quite well established that this market generally does an excellent job at inducing betting lines to accurately predict game outcomes. The debate thus centers on the accompanying notion that

simple profitable betting strategies seem to endure. Overall market efficiency predicts the demise of these strategies over time as experienced bettors exploit the profit-making opportunities. It is the conjecture of this analysis that these unexplained deviations from efficiency would be consistent with the result of point shaving in the NBA.

The first research into the efficiency of the basketball betting market performed by Camerer (1989) built upon the psychological misperception of the “hot hand,” and how individuals incorrectly judge the randomness of sequences. The cognitive psychology paper by Gilovich et al. (1985) experimentally demonstrated that while subsequent shots by college basketball players are uncorrelated, bettors tend to believe the notion that a player is more likely to make (miss) his next shot given that he has made (missed) his previous. This result was attributed to one’s tendency to reject a lengthy sequence of makes or misses as part of a random distribution, albeit that the probability of encountering such a sequence is quite high. Camerer extended this hot hand fallacy to see if the basketball betting market could accurately adjust betting lines for teams on winning or losing streaks. He aggregated game result and betting line data across three seasons for teams on winning and losing streaks. He then computed the streaking team’s probability of beating the point spread in their next game given their current streak length. This data showed that the betting market is able to accurately account for games played by teams on winning streaks, while for teams on losing streaks the probabilities of beating the point spread significantly differ from (0.5). He noted that for teams on losing streaks, these probabilities fall into the profitability range for bettors, although just barely. He concluded that for “basketball betting, the hot hand fallacy exists but it is slight,” (p.1260) as “traders can study past data and exploit such errors easily” (p.1260).

One large problem with Camerer's piece is that it started with the assumption that the hot hand is a myth and simply worked to show that the betting market does not adequately adjust for streaking teams. By only measuring the winning frequency of bets against the point spread, Camerer limited himself by his design. Brown and Sauer (1993a) addressed this issue specifically, instead offering the hypothesis untested by Camerer that simultaneous adjustments occur in a team's performance level and in market point spreads during streaks. They created a point spread pricing model to estimate the baseline expected amount by which one team would be favored over another and then tested whether this value changes based on the relative streaks of the two teams. Utilizing data over six seasons, twice that of Camerer, Brown and Sauer tested (H1) whether the hot hand is irrelevant to both the betting market and game outcomes, (H2) whether people believe in a non-existent hot hand effect as posited by Camerer, and (H3) whether point spreads take into account actual hot hand effects. Their results definitively rejected H1 that streaks are irrelevant to the betting market and to team performance. However they could not reject either H2 or H3 that either the hot hand is a myth or that recent previous performance affects immediate outcomes. Brown and Sauer convincingly demonstrated that point spreads react to streaks, increasing point spreads by 0.25 points for teams on two game winning streaks and by 0.67 points for teams on four game winning streaks. However, the reason why was inconclusive as they were unable to tease out whether the betting market was reacting to game fundamentals or to noise, which in turn was providing for the profitable strategies.

This question of fundamentals or noise was the motivation in the next paper by Brown and Sauer (1993b), retesting their three hypotheses in a more generalized form. Utilizing their point spread pricing model, they demonstrated that the explanatory power of the point spread is

quite high ($R^2 = 0.889$), however there still remains a nontrivial unpredictable component. Brown and Sauer applied this model to predict games out of sample by using data over the two teams' previous twenty games to predict its next five. Even out of sample, this model explained more than 85% of game to game variation in point spreads. Additionally, less than a quarter of the out of sample forecast errors were more than two points away from zero implying the distribution of the forecast errors were tightly concentrated about zero. From testing actual score differentials, actual point spreads, and predicted point spreads they were able to once again conclusively reject H1, that the unpredictable component is pure noise. However, their tests could not distinguish between simple and strong fundamentals comprising this unknown component. Simply put, relevant information is embedded in this unpredictable element; however, it is not strong enough to be able to be used to distinguish between close point spreads, as in 1.5 versus 2. Furthermore, while this component is small relative to its more rigorous model counterpart, Brown and Sauer found that without it, point spreads would be significantly biased, the noise component being “essential to predicting the outcome of the game itself” (p.1208).

Fortunately, one major piece missing from the above studies, the possibility of serial correlation in game to game point spreads, was tested for by Oorlog (1995) and found to be irrelevant. While the above papers tested whether by the end of the season the point spread was an accurate pricing mechanism, they did not test to see if it was an accurate pricing mechanism within the season. Brown and Sauer's positive out of sample results provide the intuition for why this eventuality was not significant, however without explicitly testing for serial correlation in the point spread between games, the analysis was not complete. By testing the regression residuals, Oorlog was able to reject autocorrelation at the 5% significance level for 51 of the 54

team seasons tested. His results express that there is no such idea as momentum for or against the point spread as the wagering public accurately takes in new information, utilizing it to set future point spreads. These results support the hypothesis that the betting market uses past information to improve future predictions. Together, these four papers arrive at similar conclusions. They demonstrate that (1) the point spread is an accurate predictor of game outcomes. Then, they show that (2) the point spread reacts to relevant information to improve its game to game forecast. Next, they reveal that (3) a mix of fundamentals and noise make up the forecast errors. Finally, they conclude that (4) the efficient market should eliminate simple profitable betting strategies as the fundamentals comprising the unpredictable component become exploited and known.

At that point the pivotal piece of literature explicitly and rigorously studying the efficiency of betting lines was authored by Gandar et al. (1998). This paper demonstrated that the betting public fundamentally consists of informed traders creating the influential market forces necessary to remove biases/inefficiencies from betting lines. Instead of taking the betting line as a singular value for a specific game as was done in the previous literature, Gandar et al. gathered data over each game's opening and closing betting lines to test the market's ability to trade to a more efficient point spread. They first noted that during the eleven hour market period between opening and closing betting lines, approximately 80% of NBA games undergo betting line changes with nearly 0.33 being greater than a point in magnitude. Thus bookmakers often respond to betting imbalances by moving the point spread, in an attempt to attract more betting on the less popular position, in order to balance the dollar amount bet on either side. The authors provided four possible reasons for betting imbalances in the market: (1) release of new

information after betting lines are announced, (2) randomness of bettor order, (3) prediction errors on the part of the betting public, and (4) prediction errors on the part of the bookmaker.

The first two possibilities were dismissed as unable to explain all of the variation between opening and closing betting lines. Information release (1) did not seem to fit as the period of market activity was too short to adequately explain all market betting line changes. Order flow randomness (2) seemed implausible as large betting line changes require continued betting on one side even as that side becomes less and less favored. The remaining two possibilities provided readily testable theories. Gandar et al. first tested the relative forecast accuracy of opening and closing betting lines and found closing betting lines to be statistically superior predictors of game outcomes. This first result supports the notion that market activity improves the ability of the betting line to project final scores, exploiting biases in bookmaker's initial betting lines. However, the actual differences between the two point spread sets were quite small: not large enough to conclude that informed bettors dominate the market. More importantly, what matters to bettors is not by how much a team beats the point spread (the relative accuracy of the betting line), but simply whether their team is able to exceed it. This led Gandar et al. into the bulk of their testing, examining the magnitude and direction of betting line changes to discover if the market driven point spread movements contain information relevant to game outcomes or if they are instead noise.

The results they found are conspicuous. For every individual positive betting line change the actual team winning proportion of bets at opening betting lines is **always** greater than (0.5). Similarly, the converse is true for negative betting line changes. Additionally, for the 2,036 games in which the betting line did not change, the probability the favored team went on to beat the point spread did not significantly differ from (0.5). Simply put, betting lines move towards

the point where each team's likelihood of covering the point spread is (0.5), effectively removing biases from opening betting lines. Furthermore, when betting lines do not move, opening betting lines are already at this point of equality. As for the magnitude of the change in the point spread, Gandar et al. discovered that the larger the betting line change the larger the distortion from (0.5) in the initial point spread. These results led to the conclusion that "in almost all instances bettors are able to move lines by a magnitude sufficient to remove opening line biases by the close of betting," (p.396–397) and that "informed traders are both present and influential in this market" (p.398). The data decisively demonstrated that bookmaker's unbiased opening betting lines are not moved through bettor activity, while bookmaker's biased opening betting lines are exploited by the informed betting public⁴. Thus, this paper concurs with Brown and Sauer's (1993b) result that bookmakers incorporate most if not all of the fundamental information into opening betting lines, however this study pushed further, conclusively showing that the betting public processes privately held fundamental information to yield improvements in prices.

In conjunction with the findings of Gandar et al., Paul and Weinbach's (2005) analysis of the NBA betting market brings this literature to a crossroads. We know that the bookmakers have strong financial incentives to set unbiased betting lines and that the betting public induces betting line adjustments by exploiting biased point spreads. Furthermore, we know that closing betting lines reflect both public and private information yielding an efficient betting market. Yet, Paul and Weinbach demonstrate using closing betting lines that while the market as a whole is efficient, simple profitable betting strategies still exist. They found that the strategy of betting double-digit underdogs rejects the null hypothesis of a fair bet, and the strategy of betting double-digit home underdogs even rejects the null hypothesis of no profitability. Given both the readily available volumes of past game data and the informed bettors present in the market

consistently trading to efficient prices, why does this market fail to exploit and eliminate this profitable betting strategy across seven seasons of data?

Paul and Weinbach suggest four different possibilities for the market's failure to accurately price games with large underdogs. The first two suggestions that (1) bettors may overestimate the relative strengths of the best teams compared with the worst teams or that (2) players or coaches may "shirk" in lopsided contests, not putting forth maximum effort, do not seem credible. It has already been shown that the betting public learns from past game results incorporating information into future prices. The simple explanation of overestimation of relative team abilities seems contrived and unlikely. As for players shirking, it may very well be true that in lopsided contests the final margin of victory decreases in the final minutes as a result of decreased effort by the winning team. However, the decrease in effort by teams winning by large margins is something for which the market should be able to account. There would have to be a systematic bias of underestimating the decrease in effort by the leading team for this possibility to explain why large underdogs tend to cover too often. Otherwise, the expected decrease in effort would yield an expected point value by which the final margin decreases, which would then be integrated into the betting lines. I plan on further examining this possibility as while the narrowing of the final score should be independent from the initial point spread solely affected by the margin as the game winds down, if players are shaving points this would not be the case.

The remaining two possibilities Paul and Weinbach present, that (3) gamblers may receive a disutility from betting on large underdogs or that (4) betting limits may restrict the ability of informed traders to discipline the market are more intriguing. It is plausible that bettors may find the idea of rooting for a team to lose by less than 18 unsavory. Note however

that this is not the observed bias towards longshots found in pari-mutuel horse race betting as the betting line is set to make the chance of each team covering the point spread equal. Thus, this suggestion purports that these informed bettors are leaving profits unclaimed, making this conclusion possible, but not as compelling. A stronger argument is that bookmaker betting limits may bind expert bettors from effectively exploiting this opportunity. These limits are put in place so that one bet cannot break the bookmaker. However, as the internet has substantially grown the size of the betting pool, bettors are able to have accounts with multiple betting sites all offering the same point spreads, implying that it is unlikely this constraint is binding. Additionally, why this constraint only affects this specific subset of games is unclear; that it would not have created market inefficiencies elsewhere is unexplained. Furthermore, in response to the four explanations listed above as well as all market inefficiency arguments, the existence of sports handicapping services provides an outlet for expert bettors to sell their information to novices. Quite popular services, they allow any bettor to follow the advice of betting industry experts. By selling expert information into the marketplace, adding to the wealth of knowledge the average bettor already has at his or her fingertips, it seems contradictory that a simple profitable betting rule could persist in the market.

One legitimate problem with the above analysis is that it rests on the assumption that the bookmakers utilize the betting line to balance the amount of money bet on each side. This assumption seems sound as it is the stated goal of the company designing the betting lines and deviation from this strategy results in financial risk to the bookmaker. However in an experiment performed by Levitt (2004), he demonstrated that for the National Football League (NFL), bookmakers may be able to increase their profits by taking positions on the outcomes of football games. His data supported the theory that bookmakers are better able to predict both the

outcomes of NFL games as well as bettors' tendencies. This predictive superiority allows bookmakers to establish betting lines inducing uneven betting favoring the less likely outcome, increasing bookmaker profits despite the increased risk. However there are some issues with whether his conclusions hold when generalized away from his experiment. First, Levitt did not use market point spread and bettor's bets data. Instead, he used data taken from a contest offered through an online sports book where participants paid an upfront entry fee, picked the outcomes of five games each week, and were paid out at the end of the season corresponding to the participants rank in the number of games he or she picked correctly. Thus, bettors did not receive any direct payoff from correctly picking a specific game, only from cumulative number of wins. Also, as the payout structure of the contest was heavily skewed towards the top finishers, in order to maximize one's expected payout, there were incentives to pick contrary to one's competitors as opposed to picking the side more likely to win. Additionally, contestants could not express the relative intensity of preference over a game's outcome as all selections received equal weighting. Furthermore, the point spreads for each game were fixed at the announced point spread each Tuesday prior to the Sunday game. The set point spread feature prevented the market from adjusting the betting line throwing all of Levitt's results into question, particularly with reference to the paper by Gandar et al.

Even if for arguments sake we believe that the assumption that the betting line is not designed to equalize the money bet on either side of the point spread for football, there are still further reasons why the assumption would hold for basketball. The major reason is that just as the two games are fundamentally different, the two betting markets are fundamentally different. Football games are concentrated predominantly on Sundays attracting many more bettors and many more betting dollars than basketball. The average NFL bettor tends to be less informed,

betting instead to spice up a game or to bet on their home team, instead of systematically betting to make consistent profits. This difference is borne out by Paul and Weinbach's observation that the betting limit for an NBA game is approximately \$2,200, less than half the \$5,500 allowed per NFL game. This possibility is also consistent with Levitt's observation that if there is a large dollar amount bet outside his sample in the opposite direction of the bets within his sample, his sub-sample conclusions would be incorrect. Since the betting line only demonstrates the median of the money bet and dollar volume on each side is the key amount, this issue could be problematic. Levitt also notes unlike NBA point spreads, NFL betting lines do not tend move, despite a week long trading period. One explanation for this difference is that the ease of scoring in football is greatly decreased as points are scored predominantly in 3s and 7s with final scores averaging in the 20s. In basketball with scores averaging in the high nineties and points scored in 1s, 2s, and 3s, it is much easier for a team to shift between covering the point spread and not covering the point spread. This ease discrepancy makes it more crucial for basketball point spreads to be accurate relative to football point spreads. Thus, despite possible evidence to the contrary for football, for basketball, it makes sense that the assumption of position neutrality on the part of the bookmaker would hold.

Returning to the central question of how the NBA betting market can be dominated by informed bettors yet sustain simple profitable betting rules, this issue can be alternatively posited to ask the origin of the informed bettors' superior abilities. Gandar et al. admit that it seems unlikely that bookmakers, armed with volumes of public data and with great monetary incentives to set accurate betting lines, incorrectly price point spreads with the frequency their analysis yielded. This contradiction signaled to Gandar et al. that informed traders either are better than the bookmakers at processing the public information into teams' probabilities of covering or they

possess superior private information. The first possibility does not seem credible as the bookmaker has at least as much if not many times more incentive to price point spreads accurately than does the bettor. Failure to price accurately and adjust betting lines exposes the bookmaker to financial risk and possibly financial ruin. However, the second option, that bettors possess superior private information over game outcomes bears strong resemblance to the illegal phenomenon of insider trading in financial markets. A heretofore unexplored aspect of the NBA betting market is whether players are using their unique superior knowledge over NBA teams and direct sphere of influence to affect game outcomes.

To instead shift the attention to the National Collegiate Athletic Association's (NCAA) basketball betting market, the possibility that collegiate basketball players were impacting their play for direct monetary gain was studied in paper by Wolfers (2006). Prior to this paper, the possibility that point shaving was the cause of similar deviations from efficiency found in the college basketball point spread betting market⁵, had never been raised. Wolfers similarly details the incentives for players to point shave, noting that the outcome which maximizes the joint surplus of the favored team's players and the gamblers is for the favored team to win but fail to cover the point spread. His analysis led to the startling conclusion that approximately 6% of games in which one team is favored by twelve points or more are influenced by point shaving. Given that games involving such strong favorites represent nearly 20% of the total population of games, his conclusions suggest about 1% of all college basketball games involve gambling related corruption.

This paper analyzes the NBA betting market in a similar style to that done by Wolfers for the NCAA betting market. By attempting to demonstrate the existence of point shaving in NBA games, I hope to resolve conflict at the heart of this literature review; how informed bettors

dominate the market, while published betting rules continue to garner profits. In addition, the data set used for this analysis includes information over in-game scores. This data is new to the current literature and through its examination, this paper pushes further and studies the manner in which games conclude; specifically looking for evidence of decreased effort after the outcome is determined yet while there is still time left to affect the final margin. In this analysis I attempt to determine the incidence of point shaving in the NBA as well as unpack if and when it is happening.

V. Data Set I: Data/Methodology

In order to collect data and create a model to test for point shaving's possible incidence, I must first return to the point shaving decision theoretic framework presented in Section III. This framework detailed the underlying incentives to engage in point shaving as well as provided three critical aspects of its accomplishment. First, players choose to reduce their effort in order to not cover the betting line as this effort reduction is easily performed while the alternative of increasing one's effort may be exceedingly difficult or impossible. Second, favorites choose to shave points as the choice by an underdog to point shave implies a commitment to lose the game. Third, strong favorites can most easily shave points, as they are exposed to the least risk by maximizing the area where the favorite wins but fails to cover the point spread. Thus, if point shaving is occurring, it is in games where one team is favored by a large margin that we would expect to find distortionary effects. In trying to detect and measure these distortionary effects, I have created two data sets. The first data set, discussed in this section and the next, is used for general tests of the NBA betting market as well as an estimation of point shaving's prevalence. The second data set, discussed in Section VII and Section VIII closely examines the structure of the end of NBA games to see how outcomes relative to the point spread change in the final few

minutes and whether the betting line has any explanatory power over the probability of the favored team covering.

This first data set contains information over 15,859 NBA regular season games played between the 1993–1994 and the 2006–2007 season. For each game, the betting line and final score were recorded as well as unique game identifiers, such as the date and the teams that played. All information was taken from the available NBA final season logs freely provided at <http://www.goldsheet.com/>. In their fiftieth year of publication, the Gold Sheet has the self-proclaimed title of “America’s No. 1 Sports handicappers for over 4 decades,” (The Gold Sheet 2007) their service being to provide statistics driven insight and edges into the sports betting world. The listed point spreads in the NBA final season logs are the consistent, agreed upon closing betting lines of the various online and casino sports books from which they draw their data. While postseason data was available in the logs, playoff games were excluded from the sample as being unnecessarily different in comparison to regular season games. In addition, there were 564 games for which there were no betting lines posted for reasons, such as uncertainty over a key player’s participation. Games played at neutral sites were rare, and for simplicity were included as home and road games for the listed home and road teams. In full, this data set was assembled to test three aspects of the point shaving story.

First, this data set contains more than twice the number of seasons and regular season games used in past literature studies. By examining this enlarged data set’s basic characteristics, I wish to discover whether past conclusions regarding the betting market’s efficiency still hold. Specifically, I use the log likelihood ratio test proposed by Even and Nobel (1992) and utilized by Paul and Weinbach (2005) to test for NBA gambling market efficiency and the profitability of individual betting rules. From the perspective of the underdog, the unrestricted log likelihood

function takes the form L^u , (3). The restricted form of the log likelihood function, L^r , is derived from substituting (0.5) for (q_{hat}) in equation (3), as statistical market efficiency dictates that ($q = 0.5$). The formula to calculate the value of the likelihood ratio statistic for the efficient market null hypothesis of ($q = 0.5$) is found in (4). If a market is found to not be *statistically* efficient, the next step would be to test for the market's *economic* efficiency, or whether betting rules are profitable. To test this profitability null hypothesis, the “eleven–for–ten” trading rule dictates that ($q = 0.524$), as bettors must win more than 52.4% of their bets to offset the bookmakers take. This break–even probability p is determined by:

$$p + (1 - p)*(-1.1) = 0 \rightarrow p = 1.1/2.1 \rightarrow p = 0.524.$$

The profitability test statistic value is calculated through the equation (5). Together, (4) and (5) test for statistical and economic NBA betting market efficiency while allowing for non–normal residuals by not imposing equal mean or median restrictions on the forecast errors.

$$L^u = n[\ln(q_{\text{hat}})] + (N - n)[\ln(1 - q_{\text{hat}})] \quad (3)$$

$$2(L^u - L^r) = 2\{n[\ln(q_{\text{hat}}) - \ln(0.5)] + (N - n)[\ln(1 - q_{\text{hat}}) - \ln(0.5)]\} \quad (4)$$

$$2(L^u - L^r) = 2\{n[\ln(q_{\text{hat}}) - \ln(0.524)] + (N - n)[\ln(1 - q_{\text{hat}}) - \ln(0.476)]\} \quad (5)$$

N – total number of observations

n – number of observations in which underdog covers

q_{hat} – observed proportion of observations where the underdog covers

The second use of this data set involves graphing kernel density functions in the mode of Figure 1, from Section III. Kernel density functions graph the probabilities of drawing specific values from a sample. The relevant statistic studied is the margin relative to the point spread, or the final margin normalized for the initial prediction. These graphs show the relative occurrence rates of betting line forecast errors. These errors are tested to see if they are symmetrically distributed about zero, indicating whether some point spread adjusted margins are more likely than others. By segmenting the data, these forecast errors are then examined to see if their

distribution and symmetry change based on the size of the betting line. The two sample Kolmogorov–Smirnov test (KS–test) is employed to test both null hypotheses; that the distributions are symmetric about zero and that empirical distributions represent the same parametrical function. The KS–test generates a D–statistic from the maximum vertical deviation between two distributions, which is used to determine if two distributions significantly differ. The KS–test is non–parametric, having the advantage of making no assumption about the distribution of the data. However, this generality comes at the cost of having low power. The KS–test and kernel density functions will visually illustrate differences in the data, providing possible intuition as to their cause.

Another way to look at the symmetry of the forecast errors is the third and final use of this data set. In a world without point shaving, the assumption of symmetric forecast errors leads to equation (6).

$$p(0 < \text{Winning Margin} < \text{Spread}) = p(\text{Spread} < \text{Winning Margin} < 2 * \text{Spread}) \quad (6)$$

This equation compares the range in which teams win but fail to cover the point spread against the identical sized range on the other side of the betting line, in which teams win and cover the point spread. If the forecast errors are found to be asymmetric from the kernel density functions and KS–tests, this technique could lead to direct estimates of the incidence of point shaving. Wolfers used this shortcut to produce his 6% estimate of point shaving’s prevalence within the college basketball betting market. I will extend his assumption to the NBA betting market to perform a similar analysis. Between these three examinations, this first data set should prove illuminative as to the nature of the NBA betting market.

VI. Data Set I: Results

This section discusses the results of the three tests specified above for this first data set. Table 1 presents the relevant summary statistics for the betting lines, final margins, and forecast errors across each year and in total for the data set. We note that while the average betting line was (-6.2), the average final margin was almost (5.7), consistent with the mean forecast error of (-0.5). While the overall mean betting line and overall mean final margin are consistent with the results of Paul and Weinbach (2005), there are striking differences in the forecast errors. Paul and Weinbach found the mean forecast error for their seven season sample, similarly defined as the final score differential minus the point spread, to instead be approximately (-3.1). It was upon this value they formed their assumption as to the non-normality of the forecast errors. Whether their data set was based from the perspective of the favored team, as was done with this analysis, is unknown and could account for this discrepancy. Regardless as to the discrepancy's cause, this analysis proceeds under Paul and Weinbach's assumption of non-normal forecast errors.

Also notable from Table 1 is that across the fourteen year sample the betting market was near perfect in equating favorite and underdog winners (7805:7822), with favorites winning 49.95% of the time. This feat of efficiency leads into Table 2, where the results of the statistical and economic efficiency tests of the NBA betting market are reported. As expected, the test does not reject the null hypothesis of market efficiency for the entire market. However, as the betting strategy becomes more restrictive, instead only betting underdogs of size X or greater, there is evidence to suggest that the statistical efficiency of the market is violated. Particularly, the strategies of betting against all favorites of 10 or more points and betting against all favorites of 12.5 or more points reject a fair bet at the 5% significance level. Why the market is able to

correct for the intervening strategies of betting against all favorites of 11.5 points and betting against all favorites of 12 points is unclear. This mystery separates this analysis' results from the results of Paul and Weinbach, whose findings were more consistent. In a world with point shaving this anomaly could be explained as the result of a discontinuity over the bettor's choice to correct the market for point shaving. If bettors were to bet underdogs of 11, 10.5, and 10 points, they would induce bookmakers to lower the point spreads. If this alteration then affects players' choices over whether to shave points, bettors could become exposed. For example, if at a point spread of 10 a player chooses to shave points, the smart bet would be on the underdog. However, if enough bets are placed on the underdog to induce a betting line change to 9 points, and the player changes his decision to point shave as a result of the new betting line, then informed bettors would be betting on the statistically less favorable outcome. This possibility could explain the ambiguity of the results.

A consistent finding with Paul and Weinbach's analysis is that none of these strategies violate economic efficiency, as win percentages do not significantly exceed 52.4%. However, in further agreement with Paul and Weinbach's results, Table 3 reports that strategies of betting big *home* underdogs not only reject the null hypothesis of a statistically efficient market, but also of an economically efficient market. Particularly, the strategy of betting all home underdogs of 12.5 points or more rejects no profitability at the 1% significance level, providing an example of a profitable betting strategy, which has persisted across a fourteen season data period. This result puts at odds Gandar et al.'s conclusion of informed traders dominating the market.

In further exploring the inconsistencies in the predictive efficiency of betting lines, the kernel density graphs and KS-tests provide illustrative evidence. For these graphs studying margin relative to the point spread (MRTS), a MRTS value of positive one indicates a game

where the favorite covered the point spread by one point, while a value of negative one indicates a game where the favorite failed to cover the point spread by one point. The kernel density function thus relates the incidence rates of the MRTS, or the forecast errors. In Figure 2, we see the incidence rates of the forecast errors for the market as a whole. It seems to approximate a normal distribution, which would be expected under symmetrical forecast errors. However, the kernel density function's peak is centered at almost two, implying that the most often seen forecast error is one in which favorites cover by two points. Furthermore, the result of the KS-test for symmetry of forecast errors about zero found in Table 4 rejects this null hypothesis at the 1% significance level. Thus, while in Table 1 it was shown that the likelihood of being to the left or to the right was zero was nearly 50%, this placement of the peak implies for the entire NBA betting market, favorites are more likely to just cover games than just not cover. This result is the diametric opposite of what we would expect to see from point shaving.

However, segmenting the data according to the size of the betting line tells a different story. The choice to partition the data into three sets, small (7), medium (8) and large point spreads (9), resulted from the profitable betting strategy found in Table 3 of betting all home underdogs of 12.5 points or more. That segmentation seemed to be a logical starting point, while the second division was chosen for symmetry purposes, as well as from basketball strategy.

$$\text{Small Spreads: Betting Line} \geq -6 \quad (7)$$

$$\text{Medium Spreads: } -6.5 \geq \text{Betting Line} \geq -12 \quad (8)$$

$$\text{Large Spreads: Betting Line} \leq -12.5 \quad (9)$$

Six points represents the expected maximum attainable points from two possessions in basketball, a signifier for whether a game is close. In Figure 3, the kernel density functions for each segmented section are graphed, displaying significant differences in the probabilities of

achieving different forecast errors with respect to the initial betting line. At first glance, it is only the forecast errors of games with small betting lines that reflect the forecast errors of the entire data set. For games of medium spread, the forecast errors approximate a normal distribution, while the forecast errors of large spreads are shifted considerably into the region in which favored teams fail to cover the point spread. Thus, while the forecast errors of games of medium spread do mirror the normal distribution as expected, forecast errors for both small and large spreads deviate; teams favored by small margins tending to just cover this margin too often, and teams favored by large margins tending to just cover this margin not often enough. Since the betting line is designed such that the direction of the forecast error should be unknowable, this result implies that the betting line may be having causative effects, game outcomes reacting to this initial forecast.

Summary statistics for these divisions by betting line are presented in Table 5, as a restatement of what the graphs have shown and previous tests have concluded; teams favored by large spreads do not cover often enough. Table 6 provides new insight into the effect of the size of the point spread. First, the null hypotheses of symmetric forecast errors are soundly rejected at the 1% and 2% levels for small and large spreads respectively, while rejected at the 10% level for medium spreads. This indicates the forecast errors are asymmetrically distributed, not only for the data set as whole, but for the individual partitions as well. Furthermore, the null hypotheses that the empirical distribution of any of the divisions represents the same underlying parametrical function of another division's empirical distribution is rejected at the 1% significance level for each hypothesis.

In further interpreting Figure 3, we know from the decision theoretical framework that games with a small initial betting line are the most difficult games in which to shave points, as

there is the greatest chance that the player will shave too much and his team will end up losing. Thus, it makes sense for the forecast errors of games with small initial betting lines not to peak to the left of zero, as that direction possibly indicates point shaving. However, instead of being centered at zero, the peak is significantly shifted approximately two points to the right of zero. This shift could indicate that underdog teams are throwing games, choosing not only to lose, but to lose by more than the point spread. This would explain the shift right of the peak within the kernel density function, as final margins would be responding to the initial point spread. However, there is a more likely explanation resulting from accepted basketball strategy. At the end of close games, defined as games where one team is leading by six points or less, it is common practice for the trailing team to foul the leading team when the leading team has the ball. This strategy does two things: first, it gives the leading team an opportunity to score two points by shooting two free throws; and second, it gives the possession of the ball to the trailing team. In the absence of this strategy, the leading team could retain possession, run out the clock, and immediately win the game. By enacting this fouling strategy, the trailing team can trade free points for ball possession and the possibility of scoring. This is a trade of an increase in mean for an increase in variance, as it often causes the leading team to extend its margin of victory, while occasionally induces the leading team to lose the game. Because the margin of victory or loss is irrelevant to the teams, this strategy has no cost if it fails, and only benefit when it succeeds. Furthermore, since games with small initial point spreads are the most likely to be close at the end, the kernel density function should be centered around a positive number, due to the increase in mean this strategy induces. That the betting market is able to equate the probabilities of covering to that of not covering as indicated in Table 5, (small favorites winning only 50.35% of the time), lends credibility to the free throw theory.

While betting lines of medium size nearly conform to the idea of symmetric forecast errors, for games with large spreads, there is a distinct left shift, centering the peak of the kernel density distribution about a negative number. These are the games with the largest window to shave points, and the distribution is shifted in the direction expected of point shaving. This shift demonstrates that for games where the initial betting line is large, favorites continue to win, but do so too often by just less than the point spread relative to winning by just more than the point spread. Specifically, the graph in Figure 3 being normalized for the betting line implies too many games end up with a final margin of victory of 12 instead of 14 when the point spread is 13 and similarly final margins of 15 occur too often instead of 17 when the point spread is 16. This is strongly suggestive of point shaving's occurrence in NBA games, as simply the magnitude of the initial betting line, supposedly an unbiased predictor of game outcomes, affects the incidence of the forecast errors. However, while a simple segmentation at two values dictated by point shaving's decision theoretic framework, resulted in semi-conclusive evidence to support point shaving's existence, to estimate point shaving's prevalence, would require an explicit structural model of player behavior.

Instead of attempting to construct such a model, the asymmetry of the forecast errors found above violates the assumption necessary for equality in equation (6). This assumption was based upon the null hypothesis of no point shaving. If by extension, we take the assumption's rejection as a rejection of this null hypothesis, then the differences derived from equation (6) can be used as estimates of point shaving's incidence. Figure 4 graphs the percentages of games falling into each region at each betting line as defined by the left and right hand sides of equation (6): $0 < \text{Winning Margin} < \text{Spread}$ vs. $\text{Spread} < \text{Winning Margin} < 2 * \text{Spread}$. The upward trend in both lines is a response to the widening of the symmetric regions about the betting line,

naturally increasing the number of game outcomes falling into both regions as the point spread rises. Initially, games in which teams win and fail to cover are outpaced by the comparison region in which teams win and cover. However, starting at betting lines of -10 and upwards, there is evidence of a reversal; the evidence becoming more dramatic at point spreads past -12.5. This reversal is further suggestive of how large favorites tend to end up winning and failing to cover, rather than winning and covering. These results would be more concrete were it not for the contradictory results for point spreads of -12 and -14.5, in which favorites more often won and covered than the comparison set. However, in aggregating across this portion of the graph from betting lines of -10 and below, including these two contradictory values, favored teams won and failed to cover the point spread 40.9% of the time, compared with only 34.5% of the time, in which they won and covered. The standard error on this aggregate estimate is approximately 1.85 percentage points, implying the difference is statistically significant. If these proportions would have been equal in a world without point shaving at approximately 37.7%, then point shaving led roughly 3% of strong favorites, who would have won and covered the point spread, to alter their play and instead win without covering. While this value denotes the proportion of games where point shaving changed whether a team covered, if we further posit that half of the strong favorites would not have covered the point spread regardless of the point shaver's altered play, then the proportion of games involving point shaving would be double this estimate, or approximately 6%. To put this percentage in perspective, across the fourteen year sample, roughly five games per season are influenced by point shaving.

This figure is by no means definitive, as it results from observed outcomes between two comparison sets with a null hypothesis of no point shaving, instead of from a rigorous model. However, it is interesting that this analysis started with a conflict of a betting market inefficiency

in a market of informed traders. Point shaving was then proposed as an explanation, which provided a framework dictating where to look for the inefficiency. Then, through observed market outcomes, the kernel density functions provided evidence in support of asymmetric forecast errors relative to the betting line. This asymmetry finally allowed for the construction of these two comparison regions and estimation of point shaving's prevalence. Armed with this 6% baseline estimate of point shaving's incidence, further investigation to ensure the veracity of this estimate was undertaken in the form of data set two.

VII. Data Set II: Data/Methodology

The methodology used to reach the estimation of point shaving's prevalence from data set one was largely influenced by Wolfer's (2006) methods. However, Bernhardt and Heston (2007) have raised concerns that the method used above may dramatically overstate point shaving's incidence. They do not discount point shaving's possible existence, but rather claim that the asymmetry Wolfers found and, by extension, I found, is instead a fundamental part of basketball. They offer the explanation that in-game basketball strategy dictates the asymmetric distribution of the forecast errors and would most certainly argue that my findings of similar asymmetries for professional basketball games support their theory. In order to test Bernhardt and Heston's hypothesis, I collected data set two.

This second data set consists of the in-game score with five minutes, four minutes, three minutes, two minutes, and one minute remaining for 6,415 games played between the 2001–2002 and the 2006–2007 seasons. The margin at each minute interval was calculated and recorded from past game “play-by-plays” freely provided with each game's box score at <http://www.nba.com/>. Data over the 838 games with unavailable past “play-by-plays” were dropped from the data set; the majority of these games take place prior to March 2001. For

games in more recent seasons with missing data, the “play-by-plays” were taken from <http://www.espn.com/>, with all “play-by-plays” compared between the two sites yielding identical results. For games resulting in overtime, the overtime period was designated as the last five minutes of the game implying a tied score with five minutes remaining. The advantage of data set two is that while data set one compared differences in observed game *outcomes* given a betting line region, data set two provides information on how a game *reached* that observed outcome. This data set was assembled to test three more aspects of the point shaving story.

First, the games are once again segmented by size of the betting line: small spreads, medium spreads, and large spreads. For each of these point spread divisions, the forecast errors are graphed in kernel density functions for each minute interval. This will visibly show how the margin relative to the point spread changes and develops across the end of the game. The KS-test is used to examine these shifts for significance and evaluate if changes in basketball strategy are responsible for differences in forecast errors. In terms of small spreads, we would expect a significant shift right during the last minute or last two minutes if the free throw theory discussed above is the driving force behind the asymmetric forecast errors. In terms of medium spreads, we should expect continued near symmetry in the game’s final minutes. While for large spreads, if players are negatively affecting their play during this ending segment of the game, we would expect a significant shift left. These three segmentations should speak to whether Bernhardt and Heston’s hypothesis of in-game basketball strategy dictates the asymmetry of the forecast errors.

Second, I use regression analysis to determine if the betting line is having causative effects. The standard model in the literature is an OLS regression of the final score on the point spread with several controlling variables. However, I eschew the OLS regression for the bulk of this section instead opting for probit analysis along the lines proposed by Gray and Gray (1997).

OLS regression studies the somewhat irrelevant question of *by how much* did a team beat the point spread, while probit analysis has the key advantage of studying simply *whether* a team beat the point spread. The explanatory variables new to this literature are the margin relative to the point spread at each minute interval. Expected to be highly correlated with each other and with the dummy independent variable for whether the favored team covered or not, the baseline regression takes the form of equation (10).

$$\textit{Favorite Covers} = \mathbf{B}_0 + \mathbf{B}_1\textit{MRTS}_{5mins} + \mathbf{B}_2\textit{MRTS}_{4mins} + \mathbf{B}_3\textit{MRTS}_{3mins} + \mathbf{B}_4\textit{MRTS}_{2mins} + \mathbf{B}_5\textit{MRTS}_{1min} \quad (10)$$

The next step is to introduce a betting line variable to discover if when controlling for the margin relative to the point spread at the end of the game, whether the size of the betting line affects the probability of covering (11).

$$\textit{Favorite Covers} = \mathbf{B}_0 + \mathbf{B}_1\textit{MRTS}_{5mins} + \mathbf{B}_2\textit{MRTS}_{4mins} + \mathbf{B}_3\textit{MRTS}_{3mins} + \mathbf{B}_4\textit{MRTS}_{2mins} + \mathbf{B}_5\textit{MRTS}_{1min} + \mathbf{B}_6\textit{Line} \quad (11)$$

As the betting line is designed to have zero explanatory power over whether a favorite covers, we would expect the betting line to continue to be uncorrelated at the end of the game. If it does have explanatory power over whether a favored team covers, given the margin by which they are covering as the game is concluding, this could be indicative of point shaving. One further step would be to once again segment the betting line to see how different betting line regions influence the probability of covering (12).

$$\textit{Favorite Covers} = \mathbf{B}_0 + \mathbf{B}_1\textit{MRTS}_{5mins} + \mathbf{B}_2\textit{MRTS}_{4mins} + \mathbf{B}_3\textit{MRTS}_{3mins} + \mathbf{B}_4\textit{MRTS}_{2mins} + \mathbf{B}_5\textit{MRTS}_{1min} + \mathbf{B}_7\textit{PointSpread}_{3.5-6} + \mathbf{B}_8\textit{PointSpread}_{6.5-9.5} + \mathbf{B}_9\textit{PointSpread}_{10plus} \quad (12)$$

These dummy variables, which indicate whether a game has a betting line in the given region, (the region of $0 \geq \text{betting line} \geq -3$ was excluded to avoid collinearity) tests if the size of the point spread affects the probability of a favorite covering. In essence, it tests to see if the betting line effects from equation (11) are non-linear.

Depending on the results of these probit regressions, a legitimate concern would be that there is correlation between the two groups of independent variables: the in-game margins relative to the point spread and the point spread. To alleviate this concern, I return to the OLS regression to test if the point spread has explanatory power over the in-game margins. This yields five equations each with the generic form of equation (13) as Y ranges from 5 to 1.

$$MRTS\ Y\ Minutes = \mathbf{B}_0 + \mathbf{B}_1 PointSpread_{0-3} + \mathbf{B}_2 PointSpread_{3.5-6} + \mathbf{B}_3 PointSpread_{10plus} \quad (13)$$

If the point spread regions are found to be uncorrelated with the in-game margins relative to the point spread, the probit results will stand.

The last aspect this data set tests, returns to the OLS regression analysis of the literature. As we will already know whether teams are more or less likely to cover the point spread given the betting line region, the OLS regression now evaluates *by how much* different point spread regions are expected to cover the point spread (14).

$$MRTS\ 0\ Minutes = \mathbf{B}_0 + \mathbf{B}_1 MRTS_{5mins} + \mathbf{B}_2 MRTS_{4mins} + \mathbf{B}_3 MRTS_{3mins} + \mathbf{B}_4 MRTS_{2mins} + \quad (14) \\ \mathbf{B}_5 MRTS_{1min} + \mathbf{B}_6 PointSpread_{3.5-6} + \mathbf{B}_7 PointSpread_{6.5-9.5} + \\ \mathbf{B}_8 PointSpread_{10plus}$$

Less relevant in the previous literature, here, equation (14) should be illuminative as to whether there is a market trade-off attempting to account for point shaving. For a given point spread level, the key comparison is the relative change in probability of covering the point spread against the expected margin by which the point spread is beaten.

Thus, this second data set tests three things. It (1) evaluates Bernhardt and Heston's hypothesis that asymmetries of forecast errors result from basketball strategy. It (2) uses probit analysis to test whether the point spread has causative effects on the probability of a team covering when controlling for margin at the end of the game. It (3) uses OLS regression to test for a market trade-off between the probability of covering and the margin by which the team covers. These analyses distinguish this paper from previous studies as this data set does not

solely have information over observed outcomes, but also data over how a game progressed to reach that market outcome. The next section discusses the results of these tests, before the last section concludes.

VIII. Data Set II: Results

To test Bernhardt and Heston's hypothesis, each betting line segmentation's forecast errors were graphed over time. Starting with small spread games, these forecast errors are shown in Figure 5, with a clear shift in the final minute. It appears evident that the two teams reach a static equilibrium in point differential slightly below the predicted point spread, which is then disrupted in response to the game's impending end. The shift moves the peak of the distribution across the zero line, implying that teams favored by small amounts widen their margin and cover the point spread in the final minute of basketball games. This widening changes the predominant forecast error from one in which favored teams are most often just failing to cover the betting line at one minute remaining to one in which favored teams are most often just covering the betting line for the final score. In testing for the significance of this density shift, the results of the KS-tests reported in Table 7 find the distribution of the final forecast errors differs at the 1% significance level when tested against every other minute's forecast errors. This difference strongly supports the free throw theory as an explanation for why small favorites tend to just cover instead of just fail to cover. The shift in basketball strategy results in a visually dramatic shift of the distribution of the forecast errors. Both the increase in the variance and the mean of the final margin precipitated by this strategy are clearly visible through the shift across the zero line in Figure 5. Thus, shifts in basketball strategy can become evident through this method.

In moving to medium spread games, the kernel density graphs in Figure 6 illustrate a similar phenomenon. While all of the forecast errors are distributed more symmetrically, there is

still a shift to the right in the final minute. Unlike small spread games, this shift induces the distribution of the final margin to be tightly centered about the betting line's prediction. Being virtually centered at zero implies the point spread provides no indication as to which side the outcome will fall among forecast errors in the immediate vicinity of the point spread. This lack of a signal is consistent with the expectations of the function of the point spread. In examining the significance of this shift, the results of the KS-tests are reported in Table 8, finding the distribution of the final forecast errors differs at the 1% significance level when tested against every other minute's forecast errors. This shift is likely a smaller version of the free throw shift found in small betting line games as fewer games are close enough to warrant it. However, as this shift both improves the predictive power of the point spread and centers the forecast errors, it is clear the betting market accounts for this smaller scale change in basketball strategy.

Finally, in examining large spread games, the kernel density graphs in Figure 7 are more ambiguous. The forecast errors are similarly distributed throughout the end of game period, providing no indication of a definite shift. However, there is a slight leftward tendency between the distribution of the final few minutes and the final score. In Figure 8, which shows only the distribution of the forecast errors at one minute and the final score, the forecast errors move out of the region where the favored team is just covering the point spread. It is unclear whether the density shift during this minute interval moves into the region where the favored team is just failing to cover the point spread, or instead moves further into the region where the favored team is covering. Regardless, Table 9 shows that the KS-test cannot even claim a shift exists between any two points within the time period. However, the KS-test's low power combined with the relatively low sample size ($n = 275$) make it difficult to determine the significance of shifts. Furthermore, point shaving's infrequency may prevent it from showing up, being subject to

additional data size limitations. Thus, while this method validated the free throw theory of trading mean for variance, it can neither speak to Bernhardt and Heston's hypothesis that basketball strategy dictates changes in probabilities of covering nor to point shaving's possible prevalence.

The results of the probit model are more definitive, with the point spread being found to have significant causative effects on the probability of the favorite covering. The point spread betting market is fundamentally based on the idea that the betting line provides no indication as to whether a team will cover. If it did provide an indication, there would be incentives to bet on one side relative to the other, implying money making opportunities for bettors to exploit, and inducing betting line changes by the bookmaker to avoid financial risk. The probit models' results, provided in Table 10, reflect these market fundamentals as the point spread has no explanatory power prior to the game. Probit Model 1 and Probit Model 2 show that in it of itself the betting line has no predictive power over the final outcome, nor does it gain predictive power by simple betting line segmentation. Simply put, prior to the game, bettors face point spreads that do not affect the outcome relative to the point spread. However, when controlling for the minute interval point-spread-normalized-margins at the end of the game, Probit Model 3 finds the betting line is a significant predictor of the probability of the favored team covering the point spread. Furthermore, its sign indicates that as the magnitude of the point spread becomes progressively larger, the favored team becoming progressively better than their opponent, the likelihood of the favorite covering decreases. Thus, for players, when controlling for the time and score as the game concludes, the betting line is a significant determinant for whether the favored team covers. In segmenting the size of the point spread, Probit Model 4 illustrates that

this decrease in the probability of covering is non-linear, as games of large spread, those of magnitude greater than 10, are most negatively affected. This is the definition of point shaving.

An interesting benefit of how the point spread is set negates a possible legitimate objection to this analysis. We know that the betting line is designed to have no correlation over whether the favored team covers. This expectation is supported by Probit Model 1 and Probit Model 2. We also know that the margin relative to the point spread at each minute interval should be highly correlated with the probability of the favored team covering. Thus, one could be concerned that the findings of Probit Model 3 and Probit Model 4 result from an underlying correlation between the point spread variables and the interval margin variables. Table 11 provides direct OLS regression estimates for whether the independent variables are correlated, finding no significant direct correlative effects. Therefore, the argument becomes that the betting line and the interval margins relative to the point spread must be both related to some other unknown variable, point shaving, which is inducing the betting line variables' significance.

Having established that the point spread influences the probability of the favored team covering the point spread, the last aspect of data set two can be examined. While normally an irrelevant question, OLS regression analysis still answers *by how much* a favored team can expect to beat the point spread. Now however, given that the betting line influences the outcome, this estimate provides insight into the betting market's tradeoff. Table 12 reports the results showing in OLS Regression Model 12 that the betting line is significantly and negatively correlated with the expected final margin, relative to the point spread. This implies that as the point spread increases in magnitude, (a move from a point spread of -10 to a point spread of -15), the favored team is expected to exceed the point spread by a greater margin. OLS Regression Model 13 reinforces the difference in expected betting-line-normalized-final-margins between

betting line regions. Teams favored by large margins are expected to exceed the point spread by almost 2 points more relative to small spread teams, and almost 1 point more relative to medium spread teams. Thus, teams favored by large margins have both the smallest probability of covering the point spread *AND* the greatest expected covering margin. This phenomenon is conveniently explained by point shaving. As the 6% prevalence rate determined from data set one indicates, the majority of teams are not engaging in point shaving. Thus, as bettors attempt to correct the market inefficiency or account for point shaving, large betting spreads are traded downward, until the most equitable split between underdogs and favorites can be reached. At that betting line level, those shaving points would still cause more teams to fail to cover the point spread in comparison with other betting size regions. Furthermore, at that betting line level, teams without point shaving players would exceed the point spread by a greater amount than would be anticipated, as the team's skill differential is greater than the point spread predicts.

Data set two has produced novel and conclusive results. For a given point differential at the end of a game, the betting line affects the probability of that game's final margin exceeding the betting line. Thus, while bettors face point spreads uncorrelated with the probability of a team covering the point spread, players' actions are influenced by the betting line to achieve different betting market outcomes. These findings are consistent with the results of data set one and provide support for the notion that point shaving exists in the NBA.

IX. Conclusions/Implications

The premise of this paper was to empirically examine the NBA betting market to test, if in spite of the large monetary compensation NBA players earn, whether there is evidence of cheating through point shaving. Generally, the informed bettors participating in the NBA betting market are efficient in allocating point spreads to reflect differing abilities between teams.

However, the asymmetric incentives, caused by the act of setting the betting line, are being exploited. Both data sets provide strong statistical evidence to suggest that this phenomenon, traditionally associated solely with unpaid college basketball players, exists in the professional ranks. The importance of this finding is that through publicizing point shaving's existence, the probability of its detection increases, correspondingly increasing its costs, and decreasing its incidence. A drawback to the method used to identify point shaving in this analysis is that it relies on small statistical anomalies becoming observable through the large sample size. An extension of this paper would be the difficult task of discovering the players responsible for shaving points. Additionally, a major point, untouched in this examination, is that point shaving is consistently attributed to players, while coaches and referees are no less principal characters in basketball games and could be equally culpable. Further research could study substitution patterns and infractions called to test if, instead, coaches or referees are willfully affecting the final margins of games. The key economic concept underlying this paper is that with the non-linear payout structure of the NBA betting market, creating a discontinuity over the value of final game margins, individuals are altering their behavior in response to these distorted incentives.

X. Tables

Table 1:
Summary Statistics of NBA Game and Betting Market Data: 1993–1994 to 2006–2007

<i>Season</i>	<i>Betting Lines</i>			<i>Final Margins</i>			<i>Forecast Errors</i>		
	<i>M</i>	<i>SD</i>	<i>Mdn</i>	<i>M</i>	<i>SD</i>	<i>Mdn</i>	<i>M</i>	<i>SD</i>	<i>Mdn</i>
93-94	-6.9796	3.9025	-6.5	6.7763	12.445	7	-0.2033	11.557	0
94-95	-6.6966	3.9131	-6.5	6.1270	12.431	7	-0.5696	11.858	-0.5
95-96	-6.6883	3.8690	-6	6.4022	11.959	7	-0.2861	11.110	0
96-97	-6.8348	3.9211	-6.5	6.5536	11.841	7	-0.2811	11.188	0
97-98	-6.9455	4.0583	-6.5	6.5974	12.045	7	-0.3481	11.317	0
98-99	-6.0883	3.6636	-5.5	5.5254	11.508	7	-0.5629	10.980	-0.5
99-00	-6.0013	3.4579	-5.5	5.6466	12.221	6	-0.3547	11.515	0.5
00-01	-5.8217	3.3295	-5.5	4.6922	12.086	6	-1.1295	11.277	-0.5
01-02	-5.8380	3.4547	-5.5	5.3586	12.385	6	-0.4794	11.858	0.5
02-03	-5.9199	3.3916	-5.5	5.4039	12.127	6	-0.5160	11.538	0
03-04	-5.6379	3.2531	-5.5	5.2338	11.417	6	-0.4041	10.947	0.5
04-05	-6.0252	3.3904	-5.5	5.4756	11.408	6	-0.5496	10.893	0
05-06	-5.7264	3.1592	-5.5	5.0252	11.787	6	-0.7012	11.322	-0.5
06-07	-5.7553	3.1817	-5.5	5.1715	12.091	5	-0.5837	11.589	-0.5
All	-6.1967	3.5971	-5.5	5.6960	12.004	6	-0.5007	11.362	0
	<i>Favorite Wins</i>	<i>Underdog Wins</i>		<i>Pushes</i>		<i>Favorite Win Percentage</i>			
	7805	7822		232		49.95%			

Note: Forecast errors are defined as (point differential + point spread), as for this data set all games were entered from the perspective of the favorite, implying all point spreads were negative. Games with point spreads of zero were entered from the perspective of the home team.

Source: Author's Calculation

Table 2:
Log Likelihood Ratio Tests of Fair Bet and No Profitability
Strategy of Betting Big Underdogs

<i>Bet Underdog of X or More:</i>	<i>Underdog Win-Loss-Push</i>	<i>Underdog Win Percentage</i>	<i>Fair Bet</i>	<i>No Profitability</i>
13	462-409-20	53.04 %	3.2270*	0.1529
12.5	572-506-20	53.06 %	4.0433**	0.2001
12	687-646-27	51.54 %	1.2613	0.3797
11.5	825-765-27	51.89 %	2.2647	0.1556
11	976-891-31	52.28 %	3.8712**	0.0082
10.5	1166-1075-31	52.03 %	3.6962*	0.1104
10	1372-1259-39	52.15 %	4.8548**	0.0575
9.5	1605-1506-39	51.59 %	3.1510*	0.7777
9	1834-1752-59	51.14 %	1.8752	
8.5	2410-2033-59	51.28 %	2.7439*	
8	2451-2311-69	51.35 %	3.4869*	
7.5	2764-2677-69	50.80 %	1.3912	
7	3095-3033-100	50.51 %	0.6273	
All Underdogs	7822-7805-232	50.05 %	0.0185	

Note: The log likelihood test statistics have a chi-square distribution with one degree of freedom.

Critical values are 2.706 ($\alpha = 0.10$), 3.841 ($\alpha = 0.05$), and 6.635 ($\alpha = 0.01$).

* is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

Table 3:
Log Likelihood Ratio Tests of Fair Bet and No Profitability
Strategy of Betting Big Home Underdogs

<i>Bet Underdog of X or More:</i>	<i>Underdog Win–Loss–Push</i>	<i>Underdog Win Percentage</i>	<i>Fair Bet</i>	<i>No Profitability</i>
13	28–11–3	71.79	7.6648***	6.1331**
12.5	39–17–3	69.64	8.8802***	6.9105***
12	48–26–3	64.86	6.6405***	4.7116**
11.5	63–40–3	61.17	5.1795**	3.2212*
11	76–57–4	57.14	2.7236*	1.2146
10.5	104–82–4	55.91	2.6083	0.9337
10	138–114–4	54.76	2.2892	0.5738
9.5	180–150–4	54.55	2.7310*	0.6209
9	221–199–7	52.62	1.1529	0.0095
8.5	289–259–7	52.74	1.6432	0.0279
8	356–312–10	53.29	2.9003*	0.2231
7.5	421–409–10	50.72	0.1735	
7	514–510–10	50.20	0.6273	
All Home Underdogs	2238–2333–65	48.96	1.9745	

Note: The log likelihood test statistics have a chi-square distribution with one degree of freedom.

Critical values are 2.706 ($\alpha = 0.10$), 3.841 ($\alpha = 0.05$), and 6.635 ($\alpha = 0.01$).

* is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

Table 4:
Kolmogorov–Smirnov Test Results: All Games

Forecast Error Symmetry About Zero			
<i>Null Hypothesis</i>	<i>n</i>	<i>p-value</i>	<i>Result</i>
$H_0: \text{kdf}(\text{MRTS}_{\text{allgames}}) = \text{kdf}(-\text{MRTS}_{\text{allgames}})$	15859	0.000***	Not Symmetric

Note: kdf = kernel density function. * is significant at 10%. ** is significant at 5%. *** is significant at 1%.
Source: Author's Calculations

Table 5:
Summary Statistics of NBA Game Data Segmented by Betting Line

	<i>Favorite Wins</i>	<i>Underdog Wins</i>	<i>Pushes</i>	<i>Favorite Win Percentage</i>
Small Spreads	4380	4319	132	50.35%
Medium Spreads	2919	2931	80	49.90%
Large Spreads	506	572	20	46.94%

Source: Author's Calculations

Table 6:
Kolmogorov–Smirnov Test Results: Games Segmented by Betting Line

Forecast Error Symmetry About Zero			
<i>Null Hypothesis</i>	<i>n</i>	<i>p-value</i>	<i>Result</i>
$H_0: \text{kdf}(\text{MRTS}_{\text{smallspread}}) = \text{kdf}(-\text{MRTS}_{\text{smallspread}})$	8831	0.000***	Not Symmetric
$H_0: \text{kdf}(\text{MRTS}_{\text{mediumspread}}) = \text{kdf}(-\text{MRTS}_{\text{mediumspread}})$	5930	0.093*	Not Symmetric
$H_0: \text{kdf}(\text{MRTS}_{\text{largespread}}) = \text{kdf}(-\text{MRTS}_{\text{largespread}})$	1098	0.014**	Not Symmetric

Forecast Error Equivalence		
<i>Null Hypothesis</i>	<i>p-value</i>	<i>Result</i>
$H_0: \text{kdf}(\text{MRTS}_{\text{smallspread}}) = \text{kdf}(\text{MRTS}_{\text{mediumspread}})$	0.000***	Not Equivalent
$H_0: \text{kdf}(\text{MRTS}_{\text{smallspread}}) = \text{kdf}(\text{MRTS}_{\text{largespread}})$	0.000***	Not Equivalent
$H_0: \text{kdf}(\text{MRTS}_{\text{mediumspread}}) = \text{kdf}(\text{MRTS}_{\text{largespread}})$	0.000***	Not Equivalent

Note: kdf = kernel density function. * is significant at 10%. ** is significant at 5%. *** is significant at 1%.
Source: Author's Calculations

Table 7:
Kolmogorov–Smirnov Test p-values: Data Set 2 – Small Spreads

Forecast Error Distribution Equivalence Across Time					
<i>n = 3067</i>	<i>4 Minutes</i>	<i>3 Minutes</i>	<i>2 Minutes</i>	<i>1 Minute</i>	<i>0 Minutes</i>
<i>5 Minutes</i>	0.892	0.721	0.474	0.439	0.000***
<i>4 Minutes</i>		0.983	0.950	0.987	0.000***
<i>3 Minutes</i>			0.999	0.999	0.000***
<i>2 Minutes</i>				0.997	0.002***
<i>1 Minute</i>					0.001***

Note: * is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculation

Table 8:
Kolmogorov–Smirnov Test p-values: Data Set 2 – Medium Spreads

Forecast Error Distribution Equivalence Across Time					
<i>n = 1887</i>	<i>4 Minutes</i>	<i>3 Minutes</i>	<i>2 Minutes</i>	<i>1 Minute</i>	<i>0 Minutes</i>
<i>5 Minutes</i>	0.742	0.277	0.311	0.294	0.000***
<i>4 Minutes</i>		0.985	0.693	0.693	0.000***
<i>3 Minutes</i>			0.996	0.989	0.001***
<i>2 Minutes</i>				0.977	0.005***
<i>1 Minute</i>					0.008***

Note: * is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

Table 9:
Kolmogorov–Smirnov Test Results: Data Set 2 – Large Spreads

Forecast Error Distribution Equivalence Across Time					
<i>n = 275</i>	<i>4 Minutes</i>	<i>3 Minutes</i>	<i>2 Minutes</i>	<i>1 Minute</i>	<i>0 Minutes</i>
<i>5 Minutes</i>	0.999	0.977	0.949	0.992	0.102
<i>4 Minutes</i>		0.788	0.949	0.992	0.153
<i>3 Minutes</i>			0.999	0.992	0.266
<i>2 Minutes</i>				0.999	0.498
<i>1 Minute</i>					0.432

Note: * is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

Table 10:
Probit Models' Results: Data Set 2

X_i	All Models: $Y = \textit{Favorite Covers}$			
	Model 1	Model 2	Model 3	Model 4
<i>Line</i>	0.00339 (0.00477)		0.04368*** (0.00906)	
<i>Point Spread</i> ₀₋₃		--		--
<i>Point Spread</i> _{3,5-6}		0.02741 (0.04123)		0.05171 (0.07635)
<i>Point Spread</i> _{6,5-9.5}		-0.01294 (0.04315)		-0.25415*** (0.08197)
<i>Point Spread</i> _{10plus}		-0.07509 (0.05427)		-0.45949*** (0.10415)
<i>MRTS</i> _{5mins}			-0.02444* (0.01315)	-0.02533* (0.01321)
<i>MRTS</i> _{4mins}			0.01877 (0.01774)	0.01841 (0.01781)
<i>MRTS</i> _{3mins}			-0.00385 (0.01811)	-0.00164 (0.01819)
<i>MRTS</i> _{2mins}			0.02015 (0.01741)	0.02018 (0.01746)
<i>MRTS</i> _{1min}			0.35262*** (0.01581)	0.35392*** (0.01587)
<i>Constant</i>	0.03674 (0.03187)	0.02057 (0.03149)	0.57904*** (0.06161)	0.43535*** (0.06016)
<i>Pseudo R</i> ²	0.0001	0.0005	0.7354	0.7369
<i>n</i>	6415	6415	6415	6415

Note: Standard Errors are reported below in parentheses.

-- denotes removed due to collinearity.

* is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

Table 11:
OLS Regression Data Set 2: Probit Model Independent Variable Correlation Results

X_i	Y_i				
	Model 6	Model 7	Model 8	Model 9	Model 10
	<i>MRTS_{5mins}</i>	<i>MRTS_{4mins}</i>	<i>MRTS_{3mins}</i>	<i>MRTS_{2mins}</i>	<i>MRTS_{1min}</i>
<i>Point Spread₀₋₃</i>	--	--	--	--	--
<i>Point Spread_{3.5-6}</i>	0.13169 (0.36844)	0.04624 (0.36869)	0.12557 (0.36979)	0.19168 (0.37050)	0.22451 (0.37000)
<i>Point Spread_{6.5-9.5}</i>	0.04745 (0.38571)	0.09218 (0.38597)	0.28679 (0.38713)	0.36906 (0.38786)	0.37483 (0.38734)
<i>Point Spread_{10plus}</i>	-0.21121 (0.48493)	-0.15405 (0.48525)	0.06742 (0.48670)	0.22128 (0.48763)	0.18236 (0.48698)
<i>Constant</i>	-0.72917*** (0.28149)	-0.61490** (0.28168)	-0.67487** (0.28252)	-0.74053*** (0.28306)	-0.71528** (0.28268)
R^2	0.0001	0.0000	0.0001	0.0001	0.0001
n	6415	6415	6415	6415	6415

Note: Standard Errors are reported below in parentheses.

-- denotes removed due to collinearity.

* is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

Table 12:
OLS Regression Data Set 2: Margin Relative to the Point Spread Analysis

X_i	All Models: $Y = MRTS_{0mins}$		
	Model 11	Model 12	Model 13
$MRTS_{5mins}$	-0.03112 (0.01995)	-0.02675 (0.01972)	-0.02690 (0.01974)
$MRTS_{4mins}$	0.03945 (0.02697)	0.04291 (0.02666)	0.04365 (0.02669)
$MRTS_{3mins}$	0.01490 (0.02699)	0.01267 (0.02668)	0.01144 (0.02671)
$MRTS_{2mins}$	0.00711 (0.02669)	-0.00033 (0.02639)	0.00111 (0.02641)
$MRTS_{1min}$	0.92834*** (0.01965)	0.92985*** (0.01943)	0.92941*** (0.01944)
<i>Line</i>		-0.16743*** (0.01362)	
$Point\ Spread_{0-3}$			--
$Point\ Spread_{3,5-6}$			0.37084*** (0.11774)
$Point\ Spread_{6,5-9.5}$			0.89115*** (0.12331)
$Point\ Spread_{10plus}$			1.7063*** (0.15503)
<i>Constant</i>	0.02203 (0.04530)	-0.95205*** (0.09101)	-0.57040*** (0.09002)
R^2	0.8988	0.9011	0.9010
n	6415	6415	6415

Note: Standard Errors are reported below in parentheses.

-- denotes removed due to collinearity.

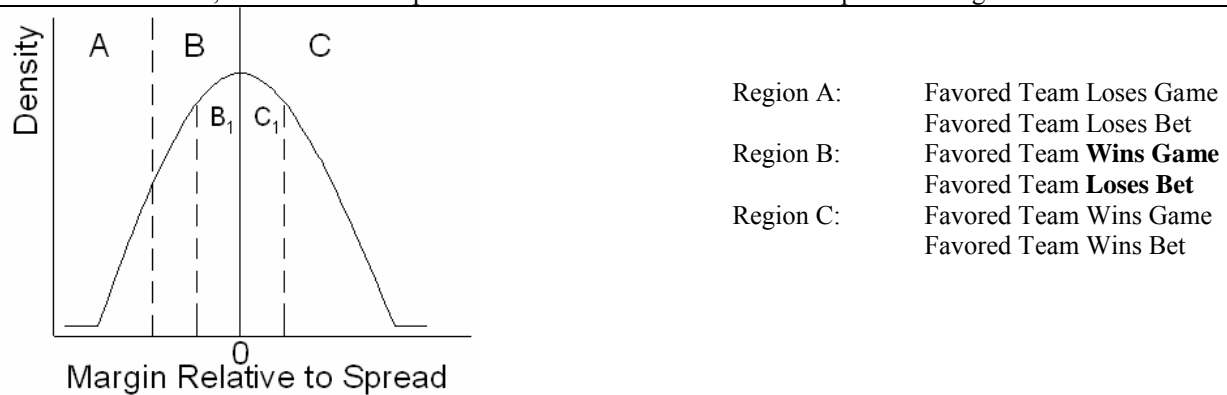
* is significant at 10%. ** is significant at 5%. *** is significant at 1%.

Source: Author's Calculations

XI. Figures

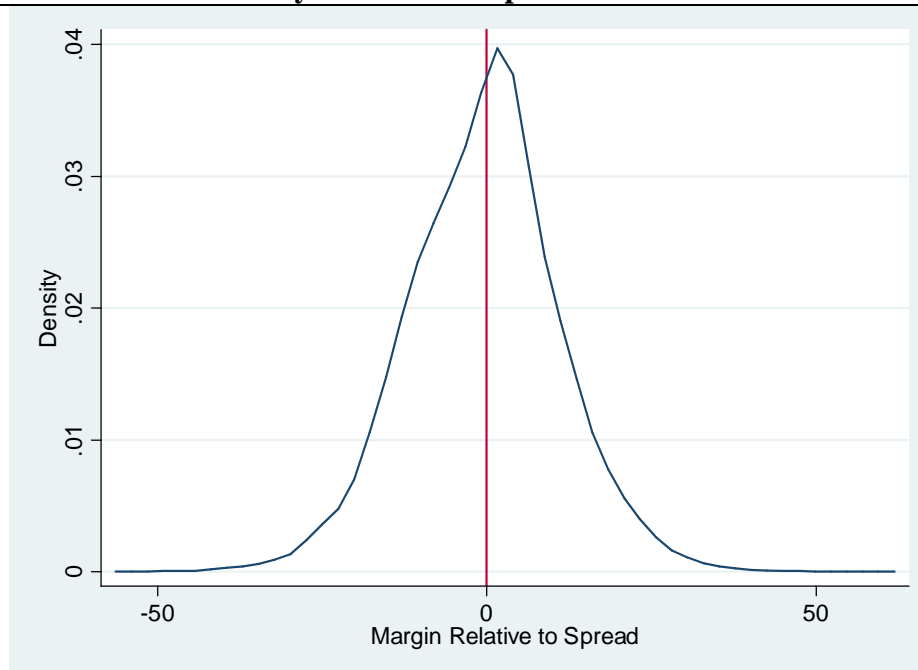
Figure 1:
Point Shaving Conceptually

The x -axis of this graph denotes the “Margin Relative to the Spread” (MRTS). A value of 0 on this axis corresponds to a game in which the favored team’s score exceeds the underdog’s score by exactly as many points as predicted by the point spread. A positive value on the x -axis therefore corresponds to a game in which the favored team covers the point spread, while a negative value corresponds to a game in which the favored team fails to cover the point spread. The point on the x -axis separating Region A and Region B corresponds to the actual score of the game. The favorite wins games to the right of this line, losing games to the left of this line. The y -axis relates the MRTS to the density of game outcomes for a given value of x . The normal distribution of games about MRTS = 0 reflects two assumptions. The first assumption is that the betting market efficiently establishes the betting line to make game outcomes equally likely on either side of the point spread. The second assumption is that the outcomes of NBA games are achieved independently from the setting of the point spread. The first assumption is addressed in the literature review; the second assumption would be violated in a world with point shaving.



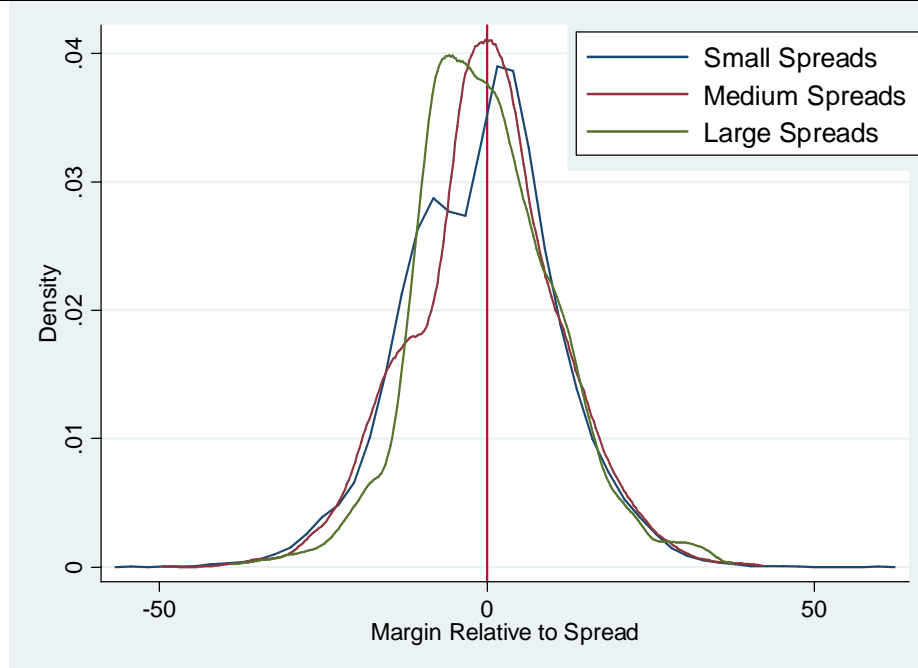
Source: Author

Figure 2
Kernel Density Function Graph of MRTS: All Games



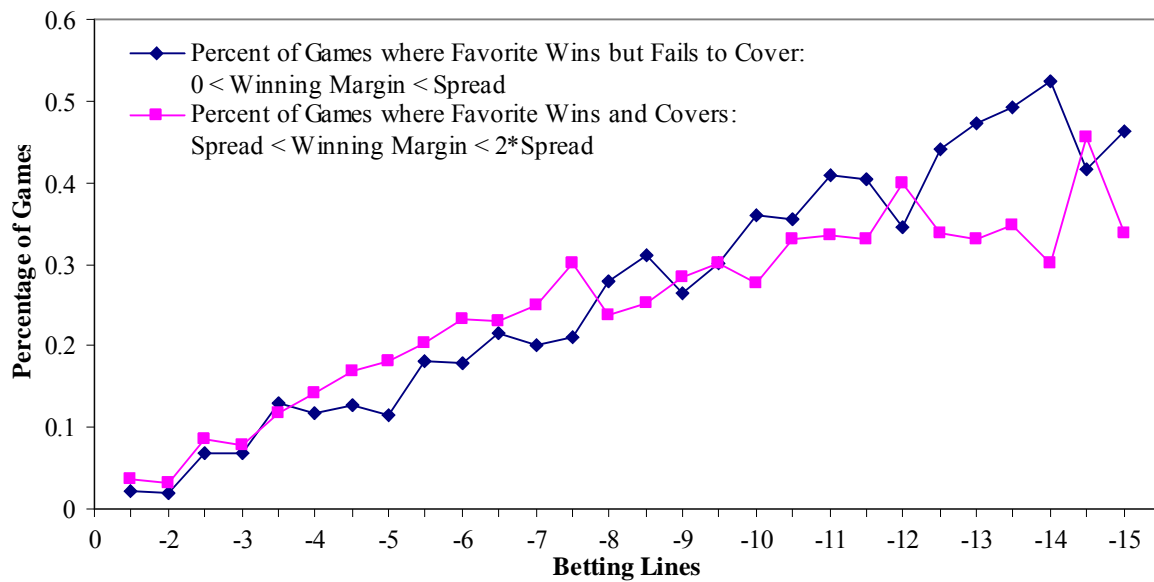
Source: Author’s Calculations

Figure 3:
Kernel Density Function Graph of MRTS: Games Segmented by Betting Line



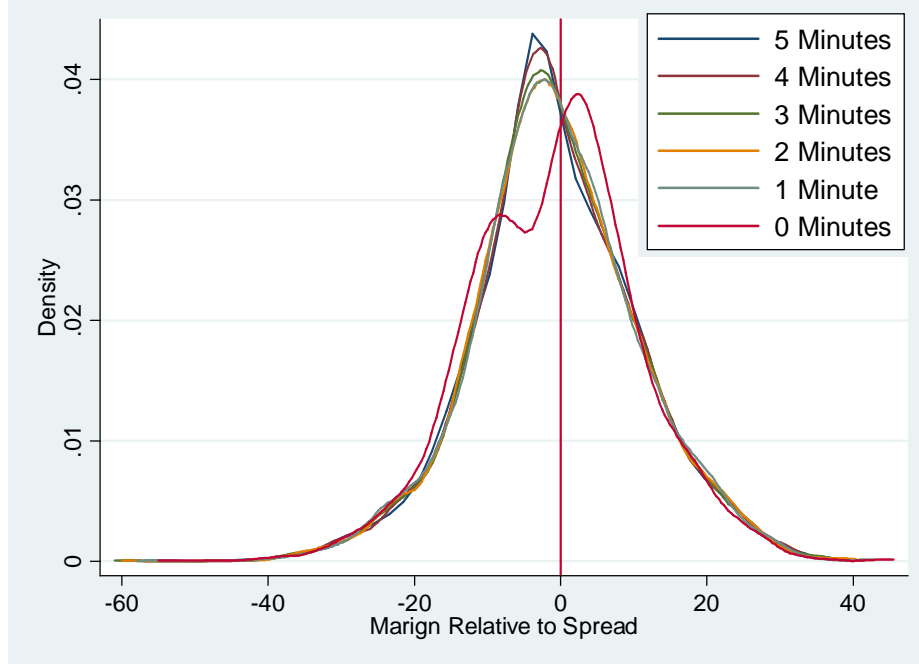
Source: Author's Calculations

Figure 4:
Betting Outcomes in Symmetric Regions about Betting Lines



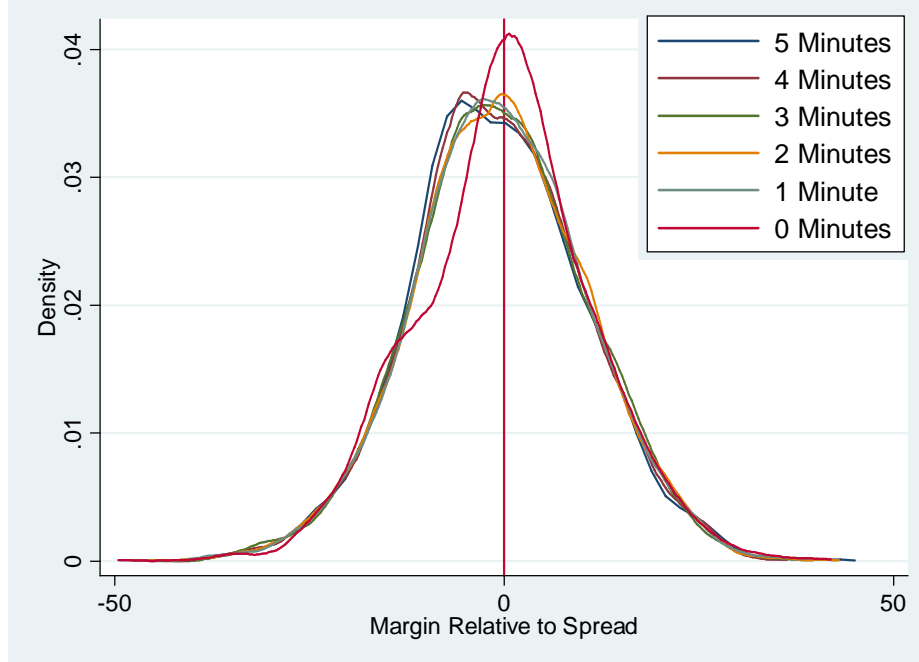
Source: Author's Calculations

Figure 5:
Kernel Density Function Graph of MRTS Across Time: Small Spreads



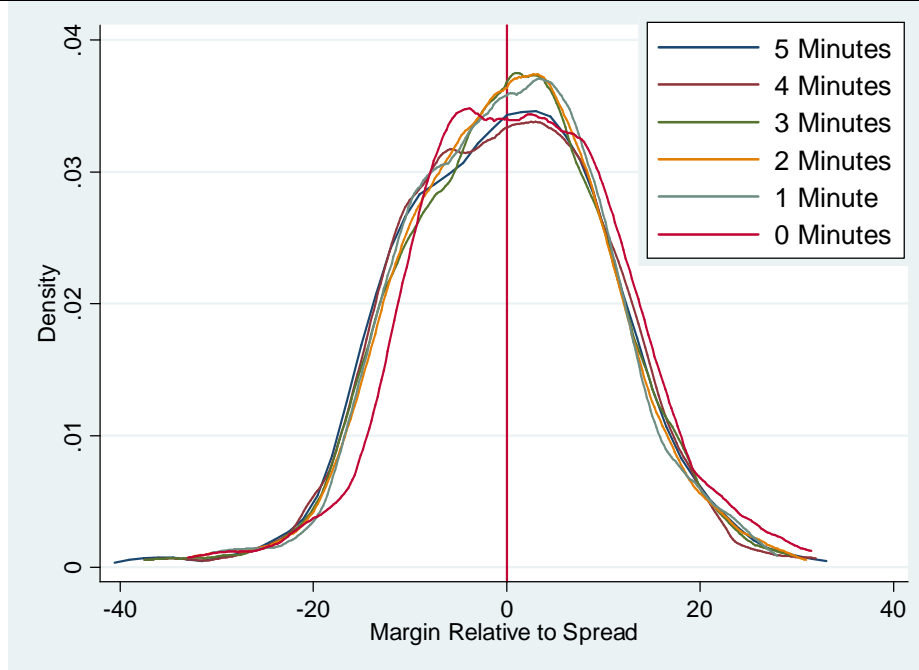
Source: Author's Calculations

Figure 6:
Kernel Density Function Graph of MRTS Across Time: Medium Spreads



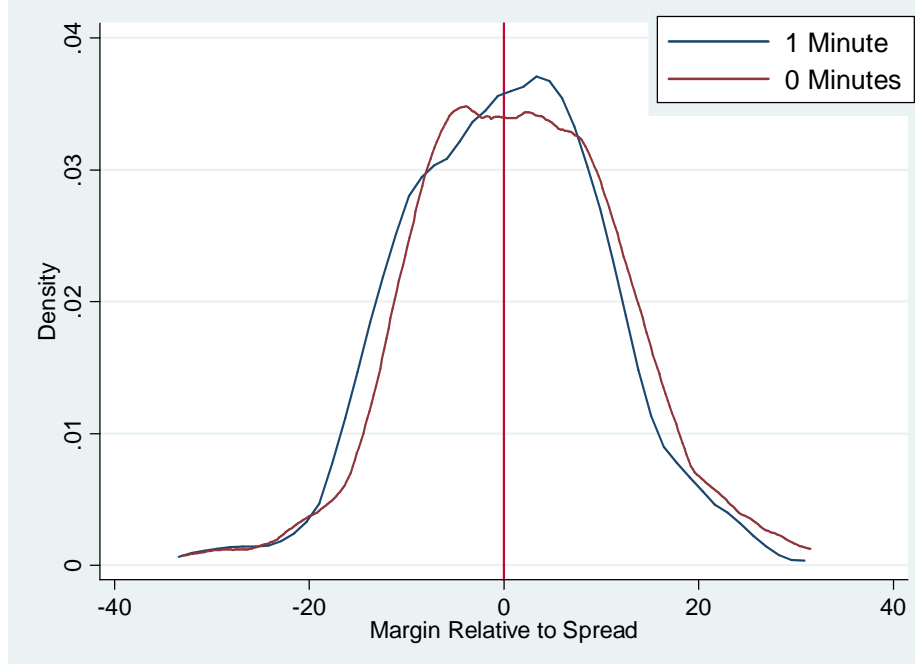
Source: Author's Calculations

Figure 7:
Kernel Density Function Graph of MRTS Across Time: Large Spreads



Source: Author's Calculations

Figure 8:
Kernel Density Function Graph of MRTS Final Minute: Large Spreads



Source: Author's Calculations

XII. Notes

1. The average NBA salary for the 1984–1985 season was \$330,000 (Ford 2006).
 $(5,200,000-330,000)/(330,000)=1475.76\%$
2. For the 2006–2007 season, the average NBA player was only 26.62 years old, averaging only 4.42 years of NBA experience. (NBA.com 2007)
3. There are volumes of stories involving NBA players and gambling. Hall of fame inductee, Charles Barkley, estimates his gambling losses at \$10 million (ESPN.com news services 2006). Journeyman and current player, Paul Shirley, names a \$500 buy-in, no-limit “hold-em” poker game as a fixture on Phoenix Suns road trips (Shirley 2005). Hall of fame inductee and current New York Knicks president of basketball operations and head coach, Isiah Thomas, used to play in a high-stakes craps game at former world champion boxer Thomas Hearns’ house (Simmons 2006). Two-time all-star and current Dallas Maverick, Jerry Stackhouse, once punched Christian Laettner, former member of the original Olympic Dream Team, during a poker game on the Detroit Pistons’ charter plane (Simmons 2006). Former all-star and fifteen year veteran, Charles Oakley, once publicly threatened former fourteen year veteran, Tyrone Hill, over a dice debt (Simmons 2006). After a practice, current all-star Gilbert Arenas bet fellow Washington Wizard teammate DeShawn Stevenson \$20,000 that Arenas could make more 3-pointers one-handed, than Stevenson could make shooting normally; Arenas won (Gilbertology.net 2007).

There are many stories involving Michael Jordan and gambling. The most famous occurred in the 1993 playoffs, as the night before a pivotal game against the New York Knicks, Jordan was seen gambling in Atlantic City (Anderson 1993). Conspiracy theorists even argue that his first retirement and subsequent two year hiatus from basketball was the result of a closed doors agreement with Commissioner David Stern in response to Jordan’s gambling. On October 6th, 1993, Jordan “retired” from basketball and three days later, the NBA announced the end of its “investigation” of Jordan's gambling finding he had violated no league rules. However, the league also admitted that its “investigation” did not even include an interview with Jordan (Kindred 2005).
4. Gandar, Zuber and Dare (2000) performed a similar study on the Totals Betting Market for NBA games and arrived at similar conclusions.

5. Colquitt et al. (2001) rigorously studied the nature of the NCAA basketball betting market. Their results are consistent with each result about the NBA basketball betting market discussed in the literature review. This consistency is found despite the added complexities of 327 Division I college basketball teams, greater asymmetric information on team abilities between geographical betting markets, and greater bettor allegiance to one's own school. This consistency in spite of the added complexity lends credence to Wolfers' examination of the NCAA basketball betting market and by extension it lends credence to applying this theory to the NBA betting market.

XIII. References

- Anderson, Dave. (1993). Jordan's Atlantic City Caper. *The New York Times* (Late Edition). May 27: p. B11.
- Associated Press. (2007). NBA scolds Arenas for making \$10 bets with fans. *ESPN.com*. March 25. <http://sports.espn.go.com/nba/news/story?id=2812075>
- Bernhardt, Dan, and Steven Heston. (2007). No Foul Play: Honesty in College Basketball. University of Illinois and University of Maryland. July.
- Brown, William O., and Raymond D. Sauer. (1993a). Does the Basketball Market Believe in the Hot Hand? Comment. *American Economic Review*, 83(5), 1377–1386.
- Brown, William O., and Raymond D. Sauer. (1993b). Fundamentals or Noise? Evidence from the Professional Basketball Betting Market. *The Journal of Finance*, 48(4), 1193–1209.
- Camerer, Colin F. (1989). Does the Basketball Market Believe in the 'Hot Hand'? *The American Economic Review*, 79(5), 1257–1261.
- Colquitt, L. Lee, Norman H. Godwin, and Steven B. Caudill. (2001). Testing Efficiency Across Markets: Evidence from the NCAA Basketball Betting Market. *Journal of Business Finance & Accounting*, 28(1), 231–248.
- Dobra, John L., Thomas F. Cargill, and Robert A. Meyer. (1990). Efficient Markets for Wagers: The Case of Professional Basketball Wagering. In Goff, Brian L., and Robert D. Tollison (Eds.), *Sportometrics* (pp. 215–249). College Station, Texas: Texas A&M University Press.
- ESPN.com news services. (2006). Barkley claims gambling problem has cost him \$10M. *ESPN.com*. May 5. <http://sports.espn.go.com/nba/news/story?id=2432043>
- Even, William E., and Nicholas R. Noble. (1992). Testing Efficiency in Gambling Markets. *Applied Economics*, 24(1), 85–88.
- Ford, Chad. (2006). Salary cap for 2006-07 season set at \$53.135M. *ESPN.com*. July 11. <http://sports.espn.go.com/nba/news/story?id=2516704>
- Gandar, John M., William H. Dare, Craig R. Brown, and Richard A. Zuber. (1998). Informed Traders and Price Variations in the Betting Market for Professional Basketball Games. *The Journal of Finance*, 53(1), 385–401.
- Gandar, John M., Richard A. Zuber, and William H. Dare. (2000). The Search for Informed Traders in the Totals Betting Market for National Basketball Association Games. *Journal of Sports Economics*, 1(2), 177–186.
- Gilbertology.net. (2007). Gilbert Arenas vs. Deshawn Stevenson. *Gilbertology.net*. February 2. <http://www.gilbertology.net/2007/02/02/gilbert-arenas-vs-deshawn-stevenson/>
- Gilovich, Thomas, Robert Vallone, and Amos Tversky. (1985). The Hot Hand in Basketball: On the Misperception of Random Sequences. *Cognitive Psychology*, 17, 295–314.
- Gray, Philip K., and Stephen F. Gray. (1997). Market Efficiency: Evidence From The NFL Sports Betting Market. *The Journal of Finance*, 52(4), 1725–1737.
- Kindred, Dave. (2005). So you want a little intrigue? Maybe Deep Throat can tell us what Michael Jordan wouldn't. *SportingNews.com*. June 24. http://findarticles.com/p/articles/mi_m1208/is_25_229/ai_n13826661
- Levitt, Steven. (2004). Why Are Gambling Markets Organised So Differently from Financial Markets?. *The Economic Journal*, 114, 223–246.
- NBA.com. (2007). 2006–2007 Player Survey: Mr. Average. *NBA.com*. http://www.nba.com/news/survey_2006.html.

- Oorlog, Dale R. (1995). Serial Correlation in the Wagering Market for Professional Basketball. *Quarterly Journal of Business and Economics*, 34(2), 96–109.
- Paul, Rodney J., and Andrew P. Weinbach. (2005). Bettor Misperceptions in the NBA: The Overbetting of Large Favorites and the “Hot Hand”. *Journal of Sports Economics*, 6(4), 390–400.
- Sheridan, Chris. (2007). Stern speaks on gambling. *ESPN.com*. February 14.
http://insider.espn.go.com/espn/blog/index?entryID=2766702&name=sheridan_chris
- Shirley, Paul. (2005). Paul Shirley’s Road Ramblings. *NBA.com*. March 19.
http://www.nba.com/suns/news/shirley_blog.html
- Simmons, Bill. (2006). Gambling and the alpha dog. *ESPN.com*. June 9.
<http://sports.espn.go.com/espn/page2/story?page=simmons/060614>
- Stern, Hal. (1999). The Man Who Makes the Odds: An Interview with “Roxy” Roxborough. *Chance*, 12, 15–21.
- The Gold Sheet. (2007). *The Gold Sheet*. <http://www.goldsheet.com/index.php>
- Wolfers, Justin. (2006). Point Shaving: Corruption in NCAA Basketball. *AEA Papers and Proceedings*, 96(2), 279–283.