MANAGEMENT SCIENCE
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0025-1909 | EISSN 1526-5501 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Inefficient Forecasts at the Sportsbook: An Analysis of Real-Time Betting Line Movement

Jay Simon
Kogod School of Business
American University
jaysimon@american.edu

This paper tests the efficiency of a set of sports betting markets using detailed betting line movement from opening until closing for four different sportsbooks for each of 3,681 Major League Baseball games. The reliability of the markets' forecasts are assessed at several lead times. The forecasts are mostly reliable, but there are simple betting strategies that would have yielded significant profit. In addition, forecasts do not always improve monotonically as the games get closer, despite more information being available; forecasts at weekend day games' start times are significantly worse than forecasts 90 minutes earlier. Furthermore, analysis of the sequences of forecasts within individual games reveals that these betting markets do not incorporate information optimally. There is sufficient evidence to reject weak form market efficiency; specifically, betting lines tend to overreact, exhibiting significant negatively autocorrelated changes that could be exploited by sophisticated bettors.

Key words: probability forecasting sequences, market efficiency, sports betting

1. Introduction

Are sports betting markets efficient? This question has been explored in depth for decades, but has generally been limited by the type of price data available. Previous analyses of sports betting markets are based on opening and/or closing prices only, primarily because that data has historically been far easier to obtain (and remains so today). This paper analyzes the sequences of available prices from four large online bookmakers from opening until closing for each of 3,681 Major League Baseball (MLB) games, yielding several new insights about how these markets behave.

The domain, while unusual, is an important one for multiple reasons. First, sports betting is an enormous industry. It is decentralized, global, and inconsistently regulated, which renders an accurate assessment of its size difficult, but two recent studies estimate its total revenue in 2020 as

Line History: Toronto Blue Jays @ Boston Red Sox

July 27, 7:10 PM

	Line (Home)	Line (Away)
07/27/21 9:21:21 AM	-130	+110
07/27/21 9:56:26 AM	-140	+120
07/27/21 1:42:10 PM	-135	+115
07/27/21 5:33:27 PM	-140	+120
07/27/21 6:44:51 PM	-145	+125

Figure 1 Moneyline history for a single example game

US$67 billion (Grand View Research 2021) and US$131 billion (Zion Market Research 2021), with the majority generated online. In the United States specifically, the Supreme Court struck down a federal ban on sports betting in 2018, and states are increasingly legalizing online sports betting to create a new source of tax revenue (Oxford Analytica 2022). In New Jersey, for instance, 93% of sports betting revenue in January 2020 (before the COVID-19 pandemic) was generated online (Can and Nichols 2022). Thus, improving our understanding of consumers' behavior in this setting is clearly a worthy endeavor for a variety of stakeholders. Second, unlike most financial markets, an asset traded in a sports betting market has an unambiguous terminal value that is realized as soon as the game is over. This makes it much easier to evaluate the accuracies of asset prices. As stated in several previous papers discussed in Section 2, sports betting markets provide a useful laboratory for studying real-world financial decisions.

Given a large dataset of detailed betting line movement, several new avenues of exploration arise. It is now possible to examine forecasting accuracy of market prices at any given lead time, as well as how those forecasts and accuracies change over time. Real-time price movement also allows for analysis of how information is incorporated into the market. In particular, it is possible to test whether the sequences of prices satisfy weak form market efficiency (i.e. that past prices do not contain additional predictive value not already captured in the current price). This is a common point of inquiry in financial markets, and was recently introduced in the realm of probability forecasting by Regnier (2018).

Note also that the vast majority of prior academic analysis of sports betting that will be discussed in Section 2 uses data generated before online sports betting was ubiquitous. Even if analogous price movement data had been available ten or twenty years ago, it is likely that market characteristics have changed in the interim due to the dramatic increase in accessibility.

The prices analyzed in this paper are *moneyline* bets. An example of the information available for a single game is shown in Figure 1. Moneylines specify the payouts for a bet; for instance, a moneyline of -130 for the home team means that a 130 bet on the home team returns a 100 payout if successful, and a moneyline of +110 for the visiting (away) team means that a 100 bet on that team returns a 110 payout if successful. Any offered moneyline bet implies a probability of the team winning; a moneyline of -130 corresponds to a probability of $130/230 = 0.565$. If a bettor believes the team's probability of winning exceeds 0.565, then a bet on that moneyline would yield a positive expected return. Bookmakers make a profit from moneyline bets on a game by offering lines on the two teams whose implied probabilities sum to slightly higher than 1 (and by setting those lines such that a team's "true" win probability is within the bounds of the two implied probabilities)[1].

Because moneylines provide probability estimates of game outcomes, analysis of them can draw on a rich literature of probability forecasting. Correspondingly, detailed moneyline movement for a game can be viewed as a sequence of probability estimates of the same outcome. Figure 2 shows an example of line movement from DraftKings, expressed as the home team's implied winning probability, for three different games. While there is ample previous work on probability forecasting, the literature on *sequences* of probability forecasts for the same event is relatively sparse. Thus, in addition to opening up new directions of sports betting analysis, the dataset used in this paper also offers a compelling real-world example of probability forecasting sequences.

In summary, analysis of the detailed MLB line movement data will reveal three key findings:

- Individual probability forecasts made by these markets are reliable and accurate in most cases, though simple profitable betting strategies did exist for this set of games.
- While the forecasts generally improve over time, forecast quality declines significantly in the few hours before day games on weekends.
- The sequences of forecasts exhibit significant inefficiencies. Changes in betting lines are negatively autocorrelated, which is a clear violation of weak form market efficiency.

The rest of the paper proceeds as follows. Section 2 presents a review of related literature. Section 3 describes the details of the dataset, along with associated technical considerations and

[1] There are two other common types of sports bet: a *point spread* bet and an *over/under* bet. A point spread bet is similar to a moneyline bet, in that it is a bet on which team will win, but instead of adjusting the prize amount, it assigns a point handicap. If the point spread is 3.5, then the favored team must win by at least 4 points (and the underdog can win, tie, or lose by up to 3 points). Point spread bets are more common than moneyline bets for higher-scoring sports, such as football and basketball. They are relatively rare for baseball. Over/Under bets differ in that they are not bets on a team; rather, they are bets on whether the total number of points scored (or "runs" scored, in baseball) will be over or under a threshold. Unlike moneyline bets, point spread and over/under bets are often symmetric. That is, a bookmaker can set the point spread or over/under threshold such that both sides of the bet are approximately equally likely to win, and can receive the same odds. Bookmakers make a profit in that scenario simply by adding a small fee to each bet.

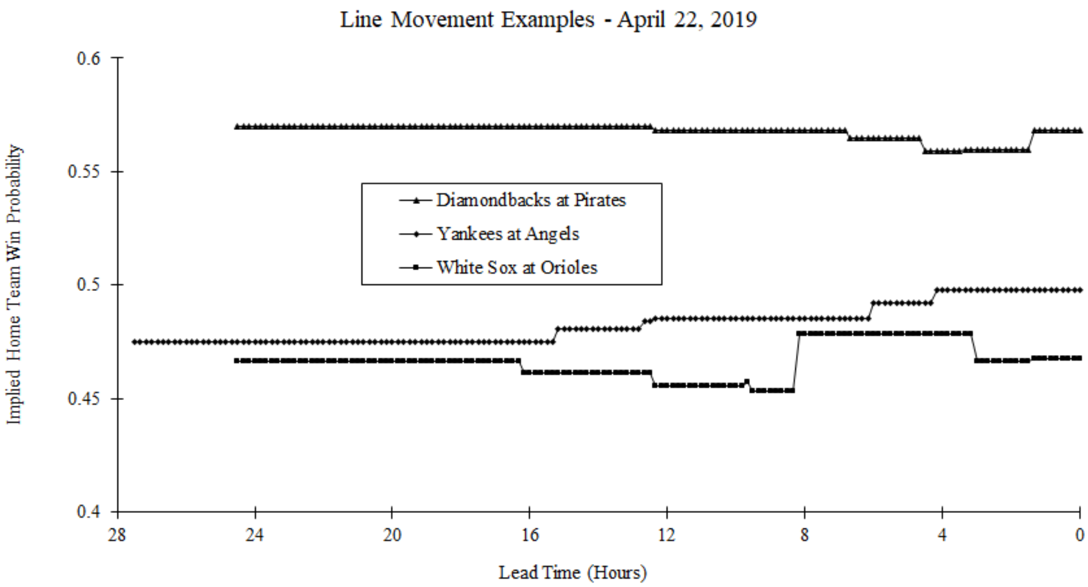

Figure 2 **Complete betting line movement (DraftKings) for three games.**

modeling choices, and presents brief comparisons between the four sportsbooks. Section 4 explores the reliability and quality of probability forecasts at individual lead times, and evaluates simple betting strategies. Section 5 examines the sequences of probability forecasts to detect possible intertemporal market inefficiencies. Finally, Section 6 concludes the paper.

2. Literature Review

The availability of sports betting data has increased steadily over the past few decades, and a rich academic literature has developed along with it. A prominent early study focused on baseball was conducted by Woodland and Woodland (1994), who analyzed closing lines from MLB games from 1979-1989, and concluded that they were efficient, in that differences between implied probabilities and actual results were not large enough to overcome fees. They also observed that betting on underdogs outperformed betting on favorites; that is, the closing lines reflected a bias toward favorites. However, Gandar et al. (2002) corrected several of the statistical calculations and observed only a bias toward slight favorites. Again, the bias was small enough such that it could not easily be exploited profitably. Paul et al. (2008) examined closing lines from 1990-2006, and also found insufficient evidence to reject the hypothesis that these markets were efficient. (The analysis was conducted as a preliminary check before using the lines to analyze competitive balance in MLB; see also Bowman et al. (2013).) In a recent broader study of opening and closing lines in MLB, NBA, NFL, and NHL games through 2013, Moskowitz (2021) found that closing lines generally overreacted; that is, an efficient line would be somewhere in between the opening line and the closing line.

There is also prior research specifically on MLB over/under betting. Brown and Abraham (2002) examined these markets from 1996-2000 and found inefficiencies that would have allowed for a simple profitable betting strategy in 1997. However, as noted by Paul and Weinbach (2004), that analysis treated the over/under bets as if they had even odds, when in fact they did not; the odds simply were not available. Bickel and Kim (2014) were able to obtain both over/under lines and corresponding odds for 2000-2007, and found that the market was efficient. (All of the over/under thresholds analyzed in these papers were closing lines.) In general, for either moneyline or over/under MLB betting markets, the consensus in the existing body of literature is that the markets are either efficient, or that any identified inefficiencies are within the bounds of transaction costs.

There is substantial prior work comparing opening and closing lines in other major sports. For instance, Gandar et al. (1998) showed that the opening lines set by bookmakers for NBA games were significantly less accurate than the closing lines. Miller Jr and Rapach (2013) included a third line for NFL games (called an *outlaw* line) in their analysis; the outlaw line was available to select bettors before betting was opened to the public. They found that efficiency improved with each of the three successive lines, but that inefficiencies remained even in closing lines. (The dataset used in that paper was from 1972; it is likely that relevant information is now more readily available to the betting public.) Dare et al. (2015) considered a particular subset of NBA games: those in which notable players were absent. They found that opening lines tended to be too favorable toward the team with the absent player, and that this bias was no longer present in the closing lines.

Several of the papers already mentioned note that there is no axiomatic reason for betting lines to be accurate predictors of game outcomes; that is, a bookmaker's objective is to maximize profit, which might not coincide perfectly with setting "correct" lines, nor with reflecting accurately the beliefs of the market. There has been additional work addressing this issue directly. Levitt (2004) found that bookmakers set betting lines for 2001-2002 NFL games in a manner that deliberately exploited known biases of the population of bettors; that is, they did not simply attempt to receive an equal volume of bets on each side. This was tested and confirmed by Paul and Weinbach (2007) for NFL games and by Paul and Weinbach (2008) for NBA games. In both pro and college football and basketball games from 2004-2010, Kain and Logan (2014) tested over/under thresholds and point spreads simultaneously. They found that while point spreads were accurate predictors of margin of victory, over/under thresholds were poor predictors of point totals. All of their possible explanations for this finding depended on bookmakers maximizing profits rather than predictive accuracy. While not specifically a study of betting markets, Morewedge et al. (2018) ran experiments with sports fans (baseball and others), and found in general that they were reluctant to bet against their favorite teams even when it would have made financial sense to do so, which suggests a potential source of bias.

It is also possible that the population of bettors might exhibit biases related to sequences of events, as such biases have been observed in financial markets. De Bondt and Thaler (1985) find weak form inefficiencies in the stock market; they observe that prior "losers" outperform prior "winners," indicating negatively autocorrelated returns. Several studies offer evidence for short-term underreaction and long-term overreaction to information (Poterba and Summers 1988, Cutler et al. 1991, Barberis et al. 1998), though the short-term time horizons in those analyses are still longer than the betting market duration for an individual baseball game. The findings of Seybert and Bloomfield (2009), who examine wishful betting and wishful thinking, suggest that short-term overreaction might occur in a sports betting market if bettors are inferring information from other bettors' behavior.

There is some literature examining sports betting line movement across multiple games. Dana and Knetter (1994) found that NFL betting lines did not react quickly enough to information from the previous week's games. Durand et al. (2021), on the other hand, found more recently that NFL betting lines exhibited overreaction to recent results. In college football, Bennett (2021) examined closing lines across consecutive games and found that when a team from a non-major conference played substantially better or worse than expected, the betting line for the team's next game did not adjust sufficiently. Sinkey and Logan (2014), on the other hand, found that college football teams that performed well relative to the point spread in their previous games tended to be overpriced, suggesting overreaction. Andrikogiannopoulou and Papakonstantinou (2018) found that soccer bettors were overly inclined to bet on teams on winning streaks (and overly disinclined to bet on teams on losing streaks). While these papers studied different types of line or betting sequences than those analyzed in the current paper, they do suggest the possibility of inefficiencies in how sports betting markets react to information, with evidence of both underreaction and overreaction.

Notably, all of the prior work discussed here is based on opening and/or closing lines. Sauer (2005) looked at in-game win probabilities in real-time, and suggested that exploring real-time changes to understand how efficient prices are reached is a promising avenue of research in sports betting markets. To my knowledge, there has been no published academic work since then analyzing actual sequences of betting lines for individual games.

The other area of literature important to the current paper is probability forecasting. Regnier (2018) provides perhaps the most directly relevant work, explicitly analyzing sequences of probability forecasts and offering a definition of efficiency in that setting. That definition is used in the current paper. It is analogous to weak form market efficiency, a common topic in finance (see the foundational work of Fama (1970)). It is also analogous to the approach used by Nordhaus (1987) for forecasts of quantities rather than probabilities. In each domain, the key feature of the sequence of prices or forecasts is that any useful information conveyed by previous values must be captured in the current value. This feature and its implications will be presented and explored more thoroughly in Section 5.

The current paper also compares the quality of probability forecasts at individual lead times. This requires the use of one or more *scoring rules* to evaluate the forecasts. Scoring rules have been studied extensively for decades, as they are useful for evaluating the quality of experts' probability estimates. Some early work on the topic includes De Finetti (1962), Winkler (1967), Murphy and Winkler (1970), and Savage (1971), and a helpful overview is given by Friedman (1983). Much of this literature focuses on distributional (i.e. non-binary) probability assessments, which tend to be more involved than scoring rules for forecasts of binary events. However, there are several common scoring rules that can be applied to either distributional or binary events; three such rules will be used in Section 4.

In many ways, the probabilities obtained via sports betting markets are more similar to probabilities from prediction markets than to probabilities stated by individual experts. In both sports betting and prediction markets, participants may choose either side of a bet on a binary event, which gives rise to the possibility of an equilibrium probability that can be viewed as a market forecast. A rich literature exists studying prediction markets, and a thorough review is beyond the scope of this paper; a few examples of influential work on the topic are Wolfers and Zitzewitz (2004), Manski (2006), Arrow et al. (2008), and Atanasov et al. (2017). However, there is an important difference between sports betting markets and prediction markets: in sports betting markets, direct exchanges between individual bettors are rare. Instead, a bookmaker sets the prices, and will effectively take one of the two sides of every bet. As discussed previously, the bookmaker might not be incentivized to set prices that reflect the best possible probability forecasts. In addition, once a bet is placed in a sports betting market, the bettor cannot easily sell it if the odds change. Secondary sports betting markets for longer-term bets do exist[2], but bets on individual games are much more difficult to resell.

3. Data

Betting line histories were obtained via Sportsbook Review's website for all regular season games in the 2019, 2021, and 2022 MLB seasons. (The 2020 season was omitted due to the COVID-19 pandemic; less than 40% of that season was played, and there was more uncertainty surrounding player availability and game postponement than in a typical season.) The line histories were extracted

[2] See, e.g., PropSwap.com.

	Opening	6 Hours	2 Hours	0 Hours
Mean	0.530	0.533	0.533	0.534
Standard Deviation	0.098	0.101	0.101	0.102
Interquartile Range	(0.457, 0.605)	(0.457, 0.607)	(0.457, 0.608)	(0.457, 0.610)
Minimum	0.219	0.231	0.230	0.230
Maximum	0.813	0.786	0.786	0.791

Table 1 Summary statistics of the home team win probabilities by lead time

from each individual game's HTML page using R. The actual betting lines are from Caesars, DraftKings, FanDuel, and PointsBet, all of which are large and widely known online sportsbooks. Each of the four sportsbooks has a line available from Sportsbook Review for most games, ranging from 78% (Caesars) to 90% (FanDuel). Unsurprisingly, the betting lines are extremely correlated across the four sportsbooks. This is partly due to arbitrage potential, and partly because online bettors can easily compare different lines and select the most favorable price for a desired bet. The actual game results were obtained from retrosheet.org.

The dataset was filtered in several ways to ensure validity and robustness of the results. First, only lines that opened at least 11 hours before the scheduled start time were considered. Later opening lines tended to be games for which there was still a major unresolved uncertainty (e.g. whether an ace starting pitcher would be available or not). Second, all canceled games, postponed games, and doubleheaders were excluded due to poor data quality; lines for these games had a high frequency of identifiable data errors (e.g. a large line movement recorded for one team and no line movement recorded for the other team, leading to implied probabilities whose sum was less than 1). Third, 17 additional lines were removed due to similar identifiable data errors. Finally, only games with lines from all four sportsbooks were included. In total, 3,681 games were used in the analysis, for a total of 14,724 betting line sequences.

For each of the sequences, the available moneyline bets were recorded at ten-minute intervals. Each of these available bets was converted to an implied home team win probability, resulting in a sequence of probability forecasts from each sportsbook for each game, and a total of slightly over two million individual forecasts. As mentioned in Section 1, the two teams' implied win probabilities will sum to slightly higher than 1, and thus will yield two slightly different estimates of the home team's win probability. The "true" market probability is somewhere in between these two numbers; the method used to estimate it will be presented in Section 4. Summary statistics of these implied probability forecasts at several different lead times are shown in Table 1. They do not differ substantively between individual sportsbooks.

Note that the mean probability forecasts are higher than 0.5. This is expected, as the home team has a slight advantage in most sports. However, it is rather unique in baseball for two reasons. First, the rules of the game are explicitly asymmetric and favor the home team somewhat. Second, there

	2019	2021	2022
Caesars	0.021	0.021	0.040
DraftKings	0.033	0.040	0.039
FanDuel	0.034	0.040	0.040
PointsBet	0.043	0.043	0.043

Table 2 Mean fee by year for each sportsbook.

is substantial heterogeneity in stadium characteristics (even the actual playing field dimensions!) that may affect individual players very differently. Over large samples, home team win probabilities have been relatively consistent at 53-55% throughout the sport's history (Swartz 2009).

Table 1 does not reveal any large changes by lead time, but a few smaller trends are noticeable. The mean home team win probability increases as the game gets closer, from 53.0% to 53.4%. The home team in fact won 53.2% of the games in the dataset. (A difference of 0.2 percentage points is approximately 0.24 standard errors of the estimate of the proportion of home team wins, and thus easily within the bounds of random variation.) The variability of home team win probabilities also increases as the game gets closer. This is not a surprise; we would expect the probabilities to deviate more from the base rate as more information becomes available.

While the purpose of the paper is not to compare sportsbooks to one another, there are a few notable differences between them that will be relevant to some of the results. The first difference involves the "fees"[3] charged in their lines; that is, the amount by which the sum of the implied win probabilities exceeds 1. Figure 3 shows the relationship between fees and the difference between home team and visiting team win probability for each sportsbook. FanDuel excepted, fees tend to be higher for the most and least balanced games. A likely explanation of why fees vary in this manner is that the relationship between odds and probability is nonlinear; while (-130, +110) and (-180, +160) appear to involve similar fees, their surplus probabilities are 0.041 and 0.027, respectively. (This is similar to the issue of misunderstanding miles-per-gallon differences explored by Larrick and Soll (2008).) If a sportsbook intends for the lines to appear similarly "fair" to casual bettors, the consequence is larger fees for balanced games. The least balanced games involve much larger numbers in moneylines, and even a casual bettor will be accustomed to seeing lines such as (-380, +300) rather than (-340, +320); the latter is very difficult for sportsbooks to offer profitably. Psychological pricing is a large field beyond the scope of this paper, but Larson (2014) provides a brief overview of common approaches, including framing effects of expressing price information in terms of two or more different numbers. Customers do not always do the relevant calculations (Estelami 2003, Drèze and Nunes 2004, Peine et al. 2009).

Caesars' fees appear to be lower than the other three, but that can be better understood when fees are separated by year. As shown in Table 2, Caesars' fees increased dramatically in 2022.

[3] Many different terms are used to refer to this fee, including *hold*, *vig*, *juice*, *cut*, *take*, and *margin*.

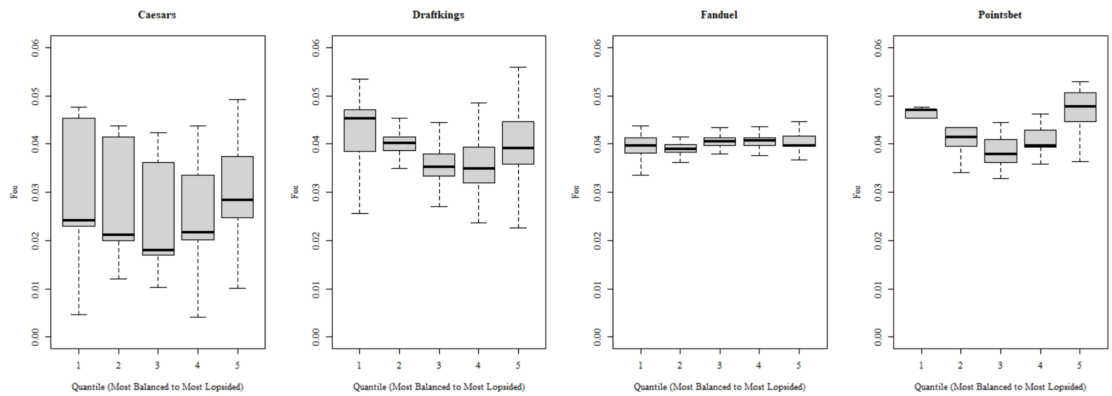

Figure 3 **Fees for each sportsbook from most to least balanced games, by quantile.**

	Caesars	DraftKings	FanDuel	PointsBet
Mean lead time (hrs.)	20.715	26.613	20.058	23.256
SD (hrs.)	4.302	7.957	4.249	4.296
Earliest open	3.6%	53.0%	4.3%	41.3%

Table 3 **Summary information of opening line lead times for the four sportsbooks**

(DraftKings' and FanDuel's fees increased somewhat in 2021 as well.) Average fees were similar for all four sportsbooks in 2022. The relationships shown in Figure 3 for each sportsbook did not change by year for Caesars, FanDuel, or PointsBet. However, DraftKings' approach in 2019 was similar to FanDuel's; they began varying fees by lopsidedness in 2021.

It is also important to note that the timing of opening lines varies substantially both by game and by sportsbook. Table 3 shows summary information of the opening lead times for each of the four sportsbooks for the 3,681 games analyzed. Opening lines are typically posted the previous afternoon or evening (US time). They often move within a few hours after being posted, but the rate of line changes is relatively low overnight, as shown in Figure 4. For individual games, later opening lines will tend to outperform the earliest opening line, as late openers can observe earlier lines and any subsequent changes. In this dataset, the earliest opening line of the four sportsbooks was posted by DraftKings or PointsBet in the vast majority of games. (In a few cases, multiple sportsbooks were denoted as earliest because they opened within the same ten-minute interval.)

4. Forecasts at Individual Lead Times

While all of the analysis in this section focuses on probability forecasts at individual lead times, it is helpful to divide the analysis into two subsections. The first subsection focuses on the reliability of those forecasts, while the second focuses on the quality of the forecasts as captured by scoring rules.

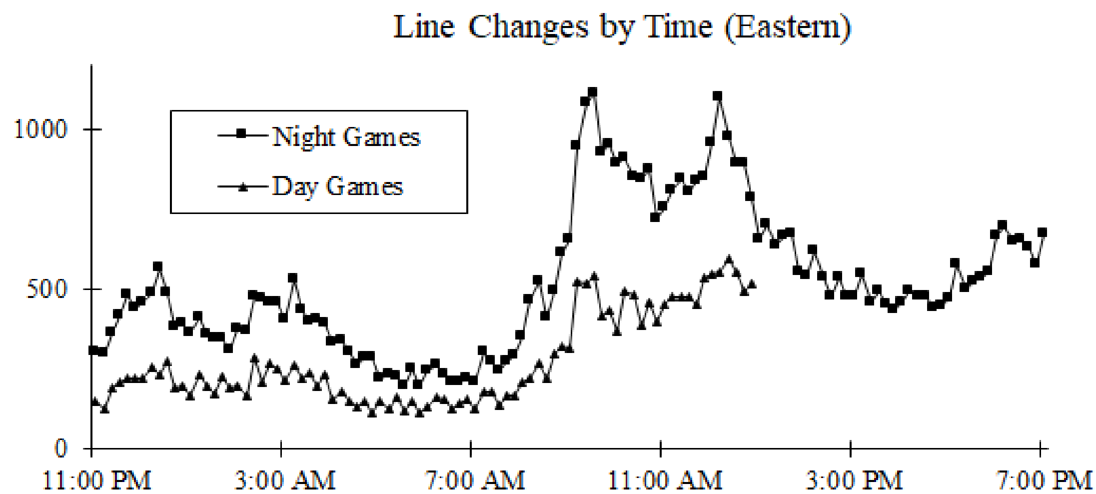

Figure 4 **Frequency of line changes by time. Only line changes within 24 hours of gametime are included. Time ranges in the chart are chosen such that most opening lines will already have been posted and games will not yet have started.**

A general point should be addressed first. As stated previously, the pairs of win probabilities (p_v, p_h) implied by moneylines will tend to have sums slightly greater than 1. Thus, either probability in isolation is a slight overestimate of the true market probability of that team winning. Two common conversion methods for obtaining the true probabilities (p_v^*, p_h^*) exist:

$$(p_v^*, p_h^*) = \left(\frac{p_v + (1 - p_h)}{2}, \frac{p_h + (1 - p_v)}{2} \right), \tag{1}$$

and:

$$(p_v^*, p_h^*) = \left(\frac{p_v}{p_v + p_h}, \frac{p_h}{p_v + p_h} \right). \tag{2}$$

The first method is used by Rascher (1999), and allocates the deduction of the "surplus" probability evenly between the two teams. The second method is used by Sauer (2005), and allocates proportional deductions. The difference between the two methods is trivial for teams that are close to evenly matched, but is much larger for lopsided games. All results shown in this paper that depend on the true market probabilities are obtained using Sauer's proportional method. (Using Rascher's equal deductions method instead leads to very similar results, likely due to the fact that opposing MLB teams tend to be evenly matched; 62% of the closing home team win probabilities are between 0.4 and 0.6.) A thorough review of these and other conversion methods is provided by Berkowitz et al. (2018).

4.1. Reliability

This subsection explores the reliability of the probability forecasts implied by the betting lines, and how it changes over time. That is, when the betting lines imply that a team has a 60% chance of winning, do those teams actually win 60% of the time?

Hypothesis 1. *Probability forecasts p_t are unreliable for at least one lead time t and observed value of p_t.*

A challenge in assessing reliability is that a given probability often does not appear frequently enough in the dataset to determine its predictive validity conclusively (though Woodland and Woodland (1994) do use that approach in a much larger dataset of closing lines). This can be resolved by splitting the probabilities into bins, and then comparing the average probability within each bin to the actual proportion of corresponding outcomes. In this case, for instance, one such bin might be $[0.55, 0.60)$. If the average forecast within that bin were 0.57, we would expect the home team to have won approximately 57% of those games.

To assess reliability, the probabilities implied by the betting lines were split into bins of width 0.05. Reliability was calculated for all bins from 0.30 to 0.75 for each sportsbook at four different lead times, as well as the opening line, and the line one hour after opening. Figure 5 shows the forecast reliability charts for PointsBet. The charts for the other sportsbooks are similar, and are shown in Appendix A. The 45-degree line reflects a perfect forecasting model. A 95% confidence interval is displayed for each bin/time pair. To determine the standard error of the sampling distributions required to construct these confidence intervals, it is necessary to incorporate the variance both in the actual outcomes and in the forecasts within each bin. The standard error of the sample proportion of wins under perfect reliability for bin k at lead time t can be calculated directly as follows:

$$\frac{\sqrt{\sum_{i=1}^{N} p_{it}(1-p_{it})X_{itk}}}{n_{tk}}, \tag{3}$$

where N is the total number of games, p_{it} is the forecast for game i at lead time t, X_{itk} is an indicator variable equal to 1 if p_{it} is in bin k and 0 otherwise, and n_{tk} is the number of forecasts at lead time t in bin k. All of the bin/time pairs in Figure 5 contained more than 100 forecasts, which is sufficiently high such that normal approximations using the calculated standard errors are reasonable. The p-values shown in Figure 5 are obtained using the test for reliability of binned forecasts given by Seillier-Moiseiwitsch and Dawid (1993); they reflect the overall reliability of that set of forecasts. The test statistic Z_t^2 is given by:

$$Z_t^2 = \sum_{k=1}^{K} \frac{n_{tk}(o_{tk} - \bar{p}_{tk})^2}{\bar{p}_{tk}(1-\bar{p}_{tk})}, \tag{4}$$

where $\bar{p}_{tk}$ is the average forecast within bin k at lead time t, and o_{tk} is the actual observed proportion of those games won by the home team. Intuitively, it is an aggregated measure of distance between the average forecast within each bin and the actual results for that bin. The null hypothesis of the test is that the forecasts are all perfectly calibrated; a large value of Z_t^2 is evidence that they are

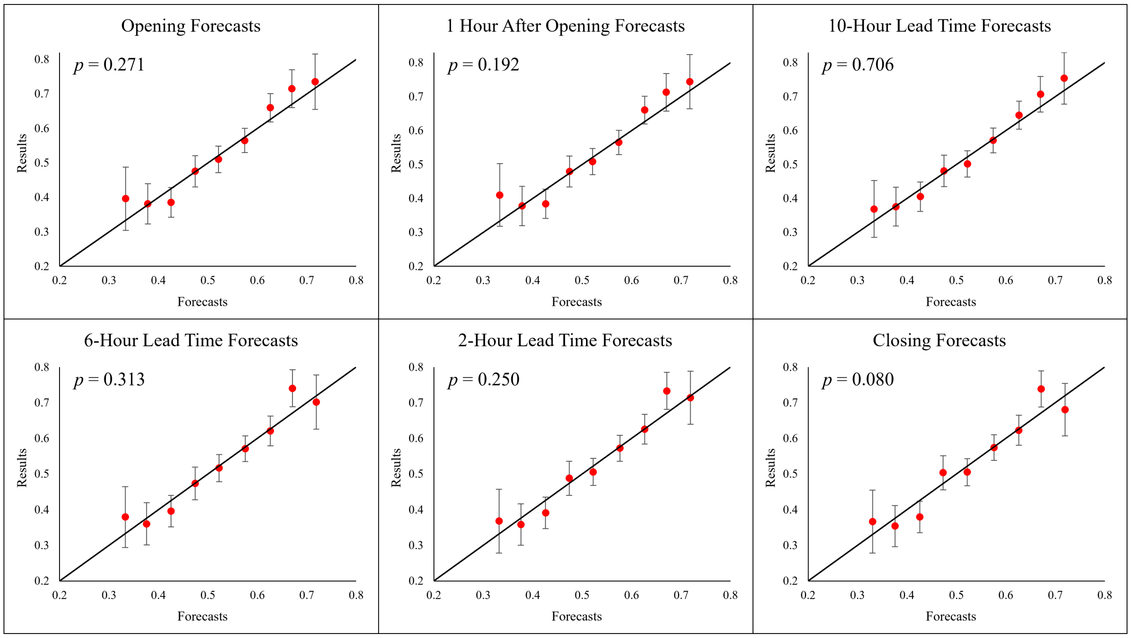

Figure 5 Reliability of PointsBet's implied home team win probability forecasts at six different times.

not. More specifically, Seillier-Moiseiwitsch and Dawid (1993) show that Z_t^2 follows a chi-squared distribution. For a helpful review of this type of statistical test, see Lahiri and Wang (2013).

These results suggest that the betting lines produced forecasts that were close to reliable, with the possible exception of some of the closing lines. The test for closing forecasts' reliability has a p-value of 0.08, as well as a few bins whose forecasts deviated noticeably from the results. (Note that the results across charts should be highly correlated, as each is illustrating predictions for the same set of games.) In summary, there is some evidence for Hypothesis 1, but not enough to establish it conclusively.

Given some modest observed unreliability, an obvious follow-up question is: does it lead to any profitable simple betting strategies? In this setting, "simple" means that the strategy can be expressed as a decision rule that depends only on the home and visiting team odds offered and the current lead time. It cannot depend on any information specific to that individual game. An example of a simple betting strategy is: "bet on home teams favored at -150 or more three hours before the game." To assess the profitability of a strategy, it is necessary both to calculate its return over the set of games and to determine whether the result is significant.

The number of possible simple betting strategies is enormous, and an exhaustive search is not practical. Results are provided for bets on home teams and visiting teams for ranges of moneyline odds at various lead times. This covers potential biases related to home teams as well as favorites and underdogs, two common issues studied by several previous papers mentioned in Section 2. The same size is used for all bets.

For each sportsbook, a total of 120 simple betting strategies were applied to the dataset, parameterized based on three factors. Half of the strategies involved bets on the home team, and half on the visiting team. Six different bet times were used: opening, one hour after opening, and lead times of ten, six, two, and zero hours. Finally, the offered home and visiting odds in the dataset were each divided into deciles[4], with strategies applied separately to each. Thus, a strategy would be expressed as, e.g.: "Bet on home teams two hours before the game on PointsBet if their moneyline odds are between -125 and -140." Because the variance of any individual strategy's return is substantial, each strategy was also run separately on the 2019, 2021, and 2022 games as a robustness check. The pairwise correlations between yearly returns for a given sportsbook range from 0.118 to 0.460, suggesting that the observed differences between strategies are not attributable solely to random noise. Unsurprisingly, returns tended to be lower in the years with larger average fees.

Each return is also tested using a null hypothesis that the implied home team win probabilities given by the sportsbook are perfectly calibrated. In the null hypothesis, the expected return for a single bet is:

$$\bar{r} = p_h^* \left(\frac{1}{p_h} - 1 \right) + (1 - p_h^*)(-1), \tag{5}$$

where p_h^* is the home team win probability implied by the pair of moneylines, and p_h is the win probability equivalent to the sportsbook's moneyline for the home team (i.e., the odds offered for the bet). The expected return of a strategy involving the home team for time t and decile d can then be expressed simply as:

$$\sum_{i=1}^{N} \bar{r}_{it} X_{itd}, \tag{6}$$

where $\bar{r}_{it}$ is the expected return of a bet on game i at time t and X_{itd} is an indicator variable similar to the one in Equation 4; it is equal to 1 if p_{it} is in decile d, and 0 otherwise. The variance of the return for a single bet is:

$$\sigma^2 = p_h^* \left(\frac{1}{p_h} - 1 - \bar{r} \right)^2 + (1 - p_h^*)(-1 - \bar{r})^2, \tag{7}$$

yielding a variance for a strategy's total return of:

$$\sum_{i=1}^{N} \sigma_{it}^2 X_{itd}, \tag{8}$$

where σ_{it}^2 is the variance of the return for a bet on game i at time t. Expected return and variance can be calculated similarly for strategies that involve betting on the visiting team. The number of bets for each strategy is sufficiently large such that the total returns are approximately normal.

[4] The deciles are of slightly different sizes. Because bookmakers offer moneylines with limited granularity, it is not possible to obtain ten equal-sized deciles.

The returns and one-tailed p-values for PointsBet are shown in Table 4; Appendix B contains the results for the other three sportsbooks, and as a further robustness check, Appendix C contains results for PointsBet when the games are split randomly into two equal-sized subsets. Variance is not constant across deciles; strategies that involve betting on longshots with occasional high payouts have higher variances than those involving bets on favorites with low payouts that usually win. Thus, statistical significance requires higher returns for the strategies that focus on longshots. The majority of returns are negative, as expected. However, several noteworthy results arise.

First, profitable strategies emerged for both home teams and visiting teams. Previous studies that looked for inefficiencies (Vergin and Sosik 1999, Schnytzer and Weinberg 2008, Wever and Aadland 2012, Shank 2018) have tended to favor betting on home teams.

Second, nearly all of the significant profitable strategies across the four sportsbooks involved betting on lopsided games. There are multiple potential explanations for this. Several of the papers discussed in Section 2 explore favorite and longshot biases; there is mixed evidence that the population of bettors could be miscalibrated on these games. It may be that it is simply easier to be well-calibrated for probabilities close to 0.5 (Lichtenstein and Fischhoff 1980). Another possible explanation is that the fee tends to be large when the teams are evenly matched for all of the sportsbooks except FanDuel, as shown in Section 3. Fees vary substantially based on how lopsided the game is. For PointsBet and DraftKings in particular, it is clearly more difficult for a bettor to make a profit betting on balanced games; see Appendix B for detailed results.

Third, significant profitable strategies tended to involve betting within the hour after opening or within the two hours before closing. The latter is particularly surprising; as explored in the next subsection, forecasts are expected to improve over time.

While the purpose of evaluating these simple betting strategies is to gain broad insights about the implication of modest deviations from reliability, there are several reasons that strategies involving particular deciles might fare unusually well or poorly. Perhaps the most obvious is a difference in fees, as discussed previously. Another is the limited granularity with which lines are set, which is further exacerbated by over-representation of round numbers in moneylines. As shown in Figure 6, frequency is not a smooth function of win probability, which suggests that there are many small distortions in the translation between "true" underlying probabilities and the lines being offered. It is possible that these distortions affect certain strategies very differently. In addition, bettors and the decision makers at these sportsbooks are human, and there are many biases related to how probabilities are expressed and interpreted. For instance, nonlinear probability weighting functions (Wu and Gonzalez 1996, Prelec 1998) could lead to lines in particular deciles being consistently too high or too low.

	Bet on home team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
$[+340, +145)$	0.056 (0.095)	0.019 (0.206)	0.012 (0.222)	0.018 (0.2)	0.03 (0.156)	0.051 (0.095)
$[+145, +120)$	-0.121 (0.928)	-0.101 (0.863)	-0.095 (0.833)	-0.106 (0.871)	-0.153 (0.973)	-0.195 (0.996)
$[+120, +104)$	-0.062 (0.668)	-0.084 (0.805)	-0.094 (0.849)	-0.119 (0.934)	-0.116 (0.922)	-0.076 (0.751)
$[+104, -110)$	-0.069 (0.653)	-0.02 (0.349)	0.03 (0.113)	0.017 (0.168)	**0.082** **(0.029)**	0.053 (0.076)
$[-110, -125)$	-0.095 (0.854)	-0.08 (0.773)	-0.046 (0.512)	-0.103 (0.887)	-0.087 (0.812)	-0.096 (0.861)
$[-125, -140)$	0.009 (0.132)	-0.012 (0.266)	-0.073 (0.775)	0.001 (0.174)	-0.013 (0.268)	-0.011 (0.258)
$[-140, -160)$	-0.041 (0.547)	-0.03 (0.445)	-0.004 (0.233)	0.013 (0.134)	-0.025 (0.403)	-0.014 (0.311)
$[-160, -180)$	-0.121 (0.97)	-0.109 (0.946)	-0.112 (0.951)	-0.109 (0.95)	-0.059 (0.682)	-0.075 (0.8)
$[-180, -220)$	**0.026** **(0.026)**	0.008 (0.075)	-0.012 (0.202)	-0.047 (0.566)	-0.069 (0.789)	-0.055 (0.658)
$[-220, -600]$	**0.012** **(0.039)**	**0.019** **(0.026)**	**0.006** **(0.048)**	**0.014** **(0.025)**	**0.009** **(0.035)**	0 (0.062)
	Bet on visiting team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
$[+435, +185)$	-0.093 (0.699)	-0.105 (0.744)	-0.067 (0.598)	-0.025 (0.406)	-0.061 (0.574)	-0.05 (0.522)
$[+185, +155)$	-0.185 (0.985)	-0.17 (0.974)	-0.16 (0.968)	-0.174 (0.981)	-0.211 (0.995)	-0.19 (0.988)
$[+155, +138)$	-0.041 (0.51)	-0.066 (0.663)	-0.03 (0.442)	0.028 (0.142)	**0.101** **(0.011)**	**0.086** **(0.021)**
$[+138, +120)$	0.032 (0.114)	0.032 (0.116)	0.008 (0.231)	-0.057 (0.632)	-0.062 (0.664)	-0.096 (0.839)
$[+120, +110)$	-0.136 (0.931)	-0.116 (0.886)	-0.117 (0.89)	-0.06 (0.631)	-0.097 (0.805)	-0.056 (0.609)
$[+110, -105)$	-0.006 (0.248)	0.002 (0.204)	0.008 (0.168)	-0.052 (0.583)	-0.025 (0.371)	-0.046 (0.535)
$[-105, -120)$	-0.027 (0.352)	-0.059 (0.614)	-0.048 (0.523)	-0.042 (0.475)	-0.048 (0.522)	-0.034 (0.413)
$[-120, -140)$	-0.064 (0.703)	-0.054 (0.62)	-0.049 (0.574)	-0.03 (0.4)	-0.004 (0.198)	0.004 (0.152)
$[-140, -168)$	**0.05** **(0.016)**	**0.065** **(0.006)**	0.024 (0.069)	**0.036** **(0.039)**	0.008 (0.143)	0.001 (0.184)
$[-168, -424]$	-0.074 (0.806)	-0.077 (0.829)	-0.057 (0.665)	-0.053 (0.62)	-0.044 (0.522)	-0.037 (0.445)

Table 4 Returns and p-values (in parentheses) for 120 simple betting strategies on PointsBet, results with p-values of 0.05 or lower bolded

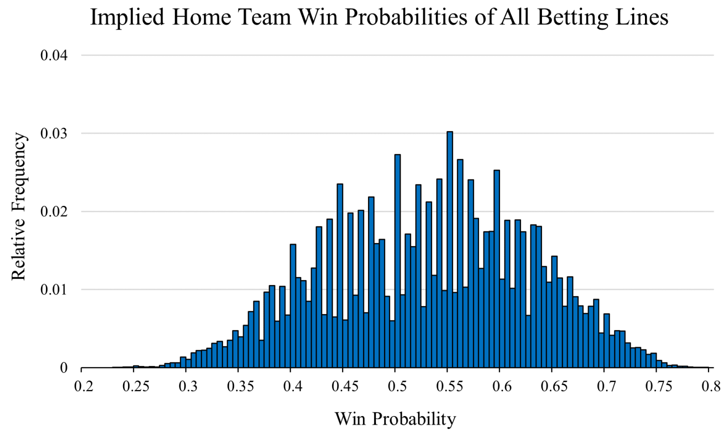

Figure 6 **Relative frequency of home team win probabilities implied by all betting lines in the dataset.**

While the p-values in Table 4 refer to the individual strategies in isolation, it is still quite clear that several of these returns are significant. If all 120 results were independent, the standard approach would be to apply a Bonferroni correction by dividing a chosen significance level by 120. However, the results are not independent. The outcome of any individual game is included in the results of 12 different strategies (via a bet on each team at six different lead times). Lines for a single game are highly correlated across lead times, but do occasionally move from one decile to another. It is not clear exactly how these dependencies should be addressed in a statistical test, but a Bonferroni or similar correction would be far more conservative than is justified, and would render identification of modest but real effects virtually impossible. Of course, regardless of the choice of statistical test, observing significantly profitable strategies does not imply that these strategies will necessarily continue to be profitable in the future. Even if positive returns cannot be dismissed as random variation, sportsbooks are capable of identifying and correcting specific sources of unreliability, and their processes are not constant over time.

4.2. Application of Scoring Rules

Even if betting lines were perfectly reliable, that would not necessarily mean they were immune to exploitation. The base rate probability of the home team winning a game in MLB was 0.532 in this dataset. In the extreme case, a bookmaker who set a line with an implied probability of 0.532 every game would be very reliable, but such a bookmaker would lose vast amounts of money. Betting lines must also diverge appropriately from the base rate using information about each individual game. That is, they should also be able to give reliable forecasts that are as far away from the base rate as possible; Regnier (2018) refers to this as *discrimination*.

There are many possible ways to evaluate the quality of probability forecasts. Proper scoring rules are preferred; i.e., a forecaster's expected performance should be optimized by stating forecasts that reflect her true beliefs. To ensure robustness, analyses for this paper were conducted using three different proper scoring rules: quadratic score (also known as Brier score, owing to Brier (1950)), logarithmic score, and spherical score. For the sake of clarity, all three of the scores are expressed such that their scales are positively oriented; i.e, higher scores are better. They are defined, respectively, as follows:

$$\begin{aligned} S_q &= 2p_r - p_r^2, \\ S_l &= ln(p_r), \\ S_s &= \frac{p_r}{\sqrt{p_r^2 + (1-p_r)^2}}, \end{aligned} \tag{9}$$

where p_r is the probability assigned to the result that occurred. Figure 7 shows the mean score for DraftKings' lines in the first three hours after opening and the final 15 hours before gametime. The charts are divided because the length of time between the line opening and the game starting varies substantially by game. As Figure 7 suggests, the choice of scoring rule does not affect any of the conclusions; hence, for brevity, all remaining scoring results will be shown using quadratic scores only[5].

Hypothesis 2. *Forecasts do not improve monotonically, as measured by a proper scoring rule, during the interval of time between opening and closing.*

Figure 7 shows that forecasts generally improve between opening and closing. However, these charts are aggregating scores for all games in the dataset, which convolutes multiple distinct patterns of performance over time. It is helpful to split the set of games into day games and night games, as betting volume by lead time differs substantially between the two. Day games typically begin between 1:00 PM and 4:40 PM ET, and night games between 7:00 PM and 10:40 PM ET. Splitting the dataset into weekday games and weekend games reveals further differences. Figure 8 shows DraftKings's scores for weekday night games and for weekend day games. (There are relatively few day games during the week and night games on weekends.) The corresponding charts for the other three sportsbooks are in Appendix C. Lead times are shown only up to ten hours for the day games, because a nontrivial proportion of these lines opened less than 15 hours before gametime.

Figure 8 reveals two very distinct patterns of market behavior. Lines for weeknight games improve during the early morning betting period and maintain roughly that level of performance until

[5] These three proper scoring rules satisfy slightly different sets of axioms. (For instance, the quadratic rule satisfies *neutrality*, meaning that the expected loss in score by stating p_1 when the "true" probability is p_2 is the same as the expected loss by stating p_2 when the true probability is p_1.) The practical differences between the scoring rules are more likely to arise when there are more than two possible outcomes, or when there are outcomes with probabilities very close to zero. A thorough comparison of various scoring rules is beyond the scope of this paper; see Selten (1998) Gneiting and Raftery (2007), or Jose et al. (2008) for details.

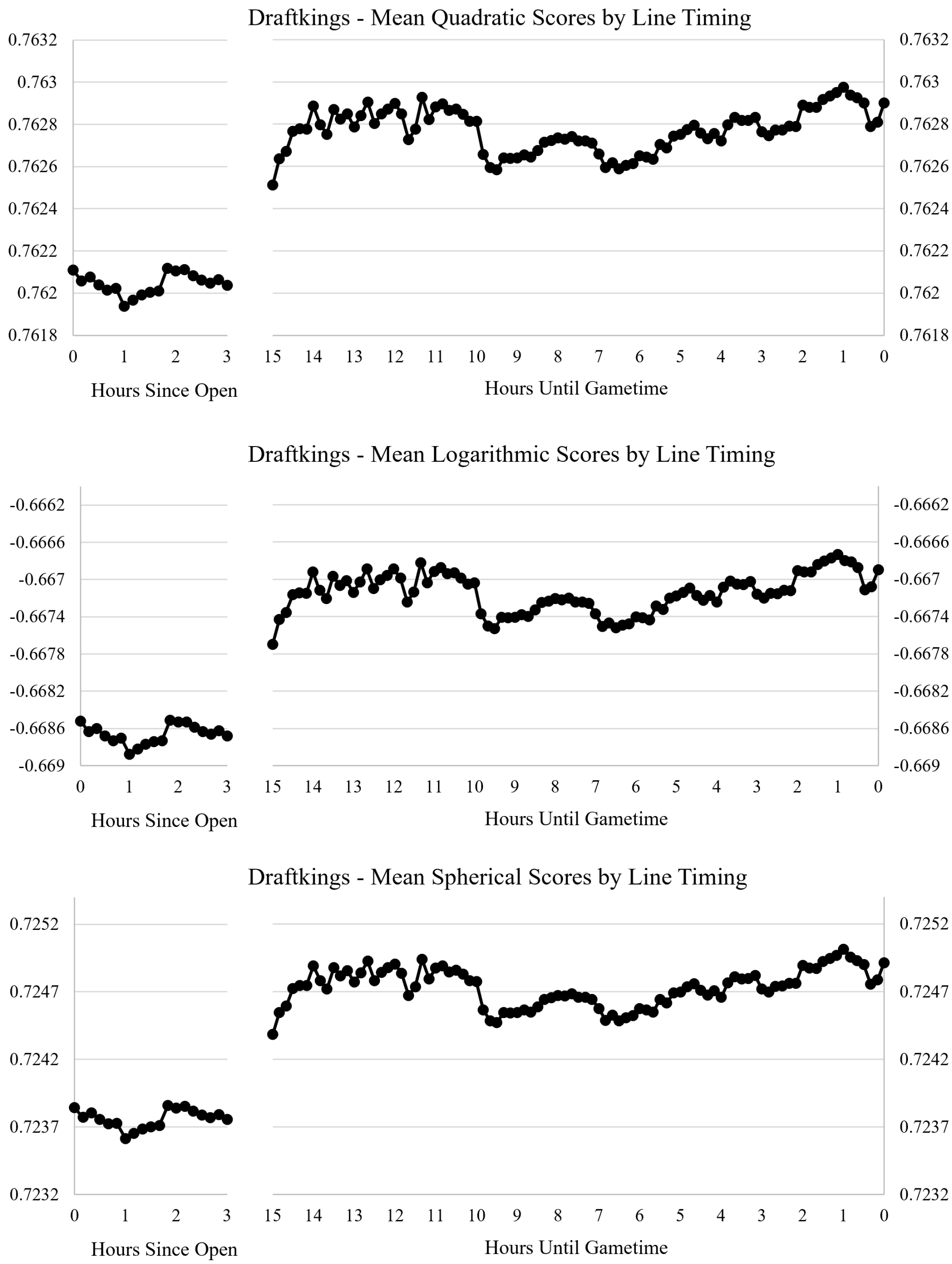

Figure 7 **Performance of DraftKings' lines using three different scoring rules.**

gametime. They do not improve several hours before the game when teams announce their starting lineups, which are a reliable source of useful information. Announced lineups are often unsurprising, but there are exceptions; for instance, a star player getting the day off unexpectedly should hurt the performance of previous forecasts relative to that of subsequent forecasts. It is possible that this type of information is often known earlier and already incorporated into prices. It is also possible

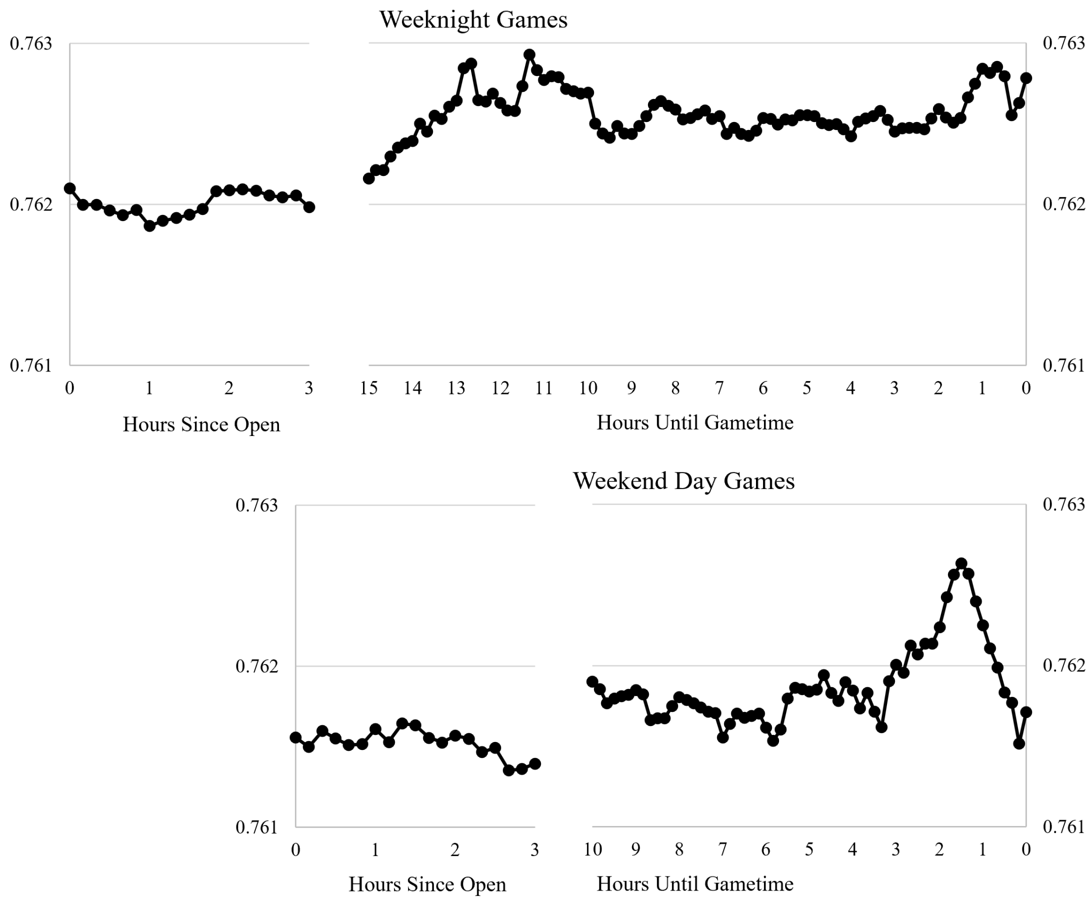

Figure 8 **Performance of DraftKings' lines for weeknight games and weekend day games.**

that it occurs rarely enough (and has a modest enough impact on win probabilities) such that its overall effect is negligible.

For weekend day games, on the other hand, Hypothesis 2 appears to be correct. It is evident by inspection that closing lines do not noticeably outperform opening lines, with a substantial decline in the final 90 minutes that erases the improvement gained during the morning and the previous night. This decline is surprising, as any efficient forecasting system should improve over time (see Proposition 4 of Regnier (2018)). The units of the y-axes in Figure 8 are in quadratic scores, whose meaning might not be intuitive. The decline in average score over the final 90 minutes can be translated approximately as follows:

There are 728 weekend day games in the dataset. In 386 of those games, DraftKings' line at gametime differed from its line 90 minutes prior. The mean magnitude of (non-zero) change in home team win probability over those 90 minutes was 0.0114. Of those 386 line changes, 213 were in the "incorrect" direction, decreasing the win probability of the team that ended up winning; the other 173 were in the "correct" direction. Thus, the 0.001 decline in mean quadratic score over 728 games is approximately equivalent to 40 line changes of slightly above one percentage point in the incorrect direction. This translation is approximate rather than exact by necessity, as quadratic score is nonlinear in probability.

To assess the significance of the decline in scores in the final 90 minutes, a linear probability model was produced with the form:

$$Y = \beta_0 + \beta_1 p_0 + \beta_2 (p_0 - p_1) + \epsilon, \tag{10}$$

where Y is the actual result of the game (1 if the home team wins, and 0 if the visiting team wins), p_0 is the closing forecast, and $(p_0 - p_1)$ is the change in the forecast during the final 90 minutes. The coefficient of interest is β_2, which is -4.496, with a p-value of 0.008. Thus, the change in market probability in the last 90 minutes leads to significantly worse forecasts. This is analogous to the approach used by Moskowitz (2021), but in this case, p_1 reflects the line 90 minutes before gametime rather than the opening line[6]. The decline in forecast quality over the final 90 minutes appears to exist for all four sportsbooks, but is only clearly significant for DraftKings and FanDuel in this dataset. (Caesars, in particular, does not move their lines as frequently, and thus many of their values of $p_0 - p_1$ are zero.)

There are several possible explanations for why this late decline in performance occurs for weekend games. One is a shift in the composition of the betting population; there are more casual bettors on weekends, and they are more likely to be placing last-minute bets[7]. Another is simply a matter of who sets the lines; while betting lines at major sportsbooks are informed by algorithms, they are ultimately set by humans, and sportsbooks often have different people making these decisions on weekends. There is also a matter of liability; due to the increase in casual bettors, betting volume is higher on weekends, which increases the likelihood of an individual decision maker moving a line to balance volume at the expense of accuracy (FanDuel 2023).

The obvious approach to constructing a simple betting strategy to exploit the decrease in forecast quality is to bet on the team whose price has decreased whenever the magnitude of line movement in the last 90 minutes before a weekend day game exceeds a threshold. Strategies using thresholds of 0.005, 0.010, 0.015, and 0.020 all yielded returns between 10 and 13 percent. Only the 0.005 threshold was significant at a 0.05 level ($p = 0.014$), though this is likely due to differences in sample size; there are 313 weekend day games in the dataset for which DraftKings' lines moved by at least 0.005 in the final 90 minutes, but only 57 such games when the threshold is 0.020.

[6] The change from opening line to closing line does not have a significant effect on win probability for the full dataset used in the current paper, nor for any of the subsets of games analyzed.

[7] Parimutuel betting markets, which are typical for horse racing, exhibit the opposite effect; early bettors are less likely to be sophisticated (Green et al. 2020).

	Caesars	DraftKings	FanDuel	PointsBet
Mean (SD) number of line changes per game	4.251 (2.517)	6.148 (3.191)	11.408 (10.267)	6.286 (3.662)
Mean (SD) size of line changes, in implied home team win probability	0.009 (0.006)	0.011 (0.008)	0.009 (0.007)	0.009 (0.005)

Table 5 Summary information of line changes for the four sportsbooks

Similar results occur for all four sportsbooks, though none are significant for Caesars, again due to a smaller number of line changes over the final 90 minutes. Of course, while these simple strategies might be less profitable in the future, it is also quite possible that this type of approach could be incorporated into a more sophisticated strategy to improve returns.

5. Sequences of Probability Forecasts

The preceding section focused on metrics describing probability forecasts at specific times, and how those forecasts changed over time in aggregate. This section will explore sequences of probability forecasts at the individual game level. Table 5 contains summary information describing line changes for each of the four sportsbooks. The number of line changes per game varies substantially by sportsbook, but the mean size of a line change (in win probability) is approximately 0.01 for each of them.

Before offering a hypothesis regarding these sequences, it is helpful to state explicitly a definition of efficiency in this setting. A system that produces sequences of probability forecasts is efficient if and only if: 1) each forecast is individually reliable, as explored in the preceding section, and 2) the sequences of probabilities are martingale; that is, $E[p_{it'}|p_{it}] = p_{it}$ for all games i and lead times t, t' such that $t > t' \geq 0$. This is the efficiency axiom provided by Regnier (2018). It is analogous to the classical result of Samuelson (1965), and is often referred to as weak form market efficiency. It implies that a probability forecast can never be improved based on knowledge of the sequence of previous probability forecasts. The results of the linear probability model in Equation (10) have already shown that efficiency does not hold for a particular subset of this dataset (efficiency implies that the coefficient of β_2 must be zero). That constitutes one specific violation of the condition, and suggests a more thorough exploration.

Hypothesis 3. *The sequences of probability forecasts are inefficient; past prices contain additional information.*

If efficiency is satisfied, previous betting lines for a game should contain no additional predictive value, and line movement should not exhibit any autocorrelation for any lag size. For one-period autocorrelation, a positive autocorrelation would indicate underreaction to information; i.e. the line not fully incorporating information immediately, and a negative autocorrelation would indicate overreaction that is subsequently corrected.

If changes to the betting line are not martingale, that will allow a sophisticated bettor to exploit them, just as a sophisticated investor could for non-martingale asset price changes in a financial market. However, there are countless potential ways in which a sequence could deviate from efficiency. Autocorrelation could be tested for any lag size, or for changes occurring at any specific pair of lead teams. These individual tests could also be aggregated to include any combination of lag sizes or pairs of lead times. While an exhaustive search would be impractical, this section presents several descriptive findings and statistical tests constituting clear evidence that betting line movement is not martingale.

The primary challenge in testing for autocorrelation in this dataset is that the number of line changes for any individual game is small, as shown in Table 5. Thus, it is difficult to observe meaningful autocorrelation for a single game. Unfortunately, there exists very little precedent for testing autocorrelation of a system that generates many time series. However, if autocorrelation metrics are computed for each game individually and each game is treated as an independent observation, then aggregating those metrics is fairly straightforward; this is the general approach taken by Regnier (2018). Because there is no single established method for detecting autocorrelation across many time series, aggregated results based on a few different autocorrelation measures are presented.

One way to measure autocorrelation for an individual time series is via an autoregressive (AR) model. An AR(m) model predicts each change based on a linear combination of the preceding m changes, producing a coefficient for each lag size $1,\ldots,m$. A coefficient of 0 indicates no autocorrelation for that lag size. In the simplest case where $m=1$, the coefficient of the AR(1) model is a measure of one-period autocorrelation. The use of autoregressive models to detect inefficiency is well-established in analysis of financial markets; see, e.g., Summers (1986), Conrad and Kaul (1988), and Fama and French (1988).

An AR(1) model was created for each sportsbook for each of the 3,681 betting lines with movement before the last period. (Movement before the last period is necessary to produce at least one non-zero value of the lagged time series.) The mean coefficients of the lagged values are shown in Table 6. If the individual coefficients are treated as independent, then the sample means for DraftKings, FanDuel, and PointsBet have two-tailed p-values that are effectively zero. It is not surprising that FanDuel's lines exhibit the strongest negative autocorrelation; their lines move much more frequently, with comparably-sized changes and no noticeable detriment to their overall accuracy. Caesars is the only sportsbook without significant autocorrelation, but it should be noted that their line changes were significantly negatively autocorrelated in 2022, suggesting that their policy changes went beyond simply an increase in fees.

Caesars	DraftKings	FanDuel	PointsBet
0.001	-0.019	-0.063	-0.027
(0.626)	(0.000)	(0.000)	(0.000)

Table 6 **Mean AR(1) coefficients for betting line changes, with p-values in parentheses.**

Length of Time Period	Caesars	DraftKings	FanDuel	PointsBet
20 minutes	-0.012	-0.032	-0.086	-0.040
30 minutes	-0.023	-0.044	-0.096	-0.050
40 minutes	-0.031	-0.054	-0.105	-0.067
50 minutes	-0.040	-0.064	-0.112	-0.074
60 minutes	-0.043	-0.067	-0.122	-0.081

Table 7 **Mean AR(1) coefficients for betting line changes with 20-60 minute period lengths (all results significant with p <0.001).**

The AR(1) results do not differ substantively for day and night games, nor for weekday and weekend games. In addition, they are robust to the exclusion of small line changes, as well as to the exclusion of portions of the interval between opening and closing. Given these AR(1) results, it is evident that Hypothesis 3 is correct for at least three of the four sportsbooks.

However, the results are likely understating the degree of autocorrelation, due to the overwhelming majority of ten-minute periods in which the line did not move. Even stronger results from autoregressive models can be obtained by aggregating the probability forecast changes into longer periods. This leads to sequences with fewer values but greater proportions of non-zero values. Table 7 shows the AR(1) test statistics for 20, 30, 40, 50, and 60 minute periods[8]. As the changes are aggregated into increasingly larger blocks of time, their negative autocorrelation becomes even more pronounced, and is clearly significant for all four sportsbooks.

To illustrate the distribution of the AR(1) coefficients, a histogram is shown in Figure 9 using 60 minute periods for FanDuel as an example. While the majority of the coefficients are close to zero, it is clear that negative autocorrelation is much more prevalent than positive autocorrelation.

Another common method for detecting autocorrelation is the variance ratio test of Lo and MacKinlay (1988). Like the use of AR(1) models, it is well-established in financial analysis; see, e.g., Liu and He (1991), Huang (1995), Smith and Ryoo (2003), and Charles and Darné (2009). The variance ratio test is based on the premise that if a time series is a homoskedastic martingale, then the variance of $(p_{i-q} - p_i)$ is q times the variance of $(p_{i-1} - p_i)$ for any lag q; that is, the variance of the change in probability from advancing q periods closer to the game's start time is q times the variance when advancing one period. (There is a heteroskedastic version as well.) The variance ratio is given by:

$$VR(q) = \frac{\sigma_q^2}{q\sigma_1^2}, \tag{11}$$

[8] When the interval from opening to closing was not divisible by the period length, a shorter first period was used.

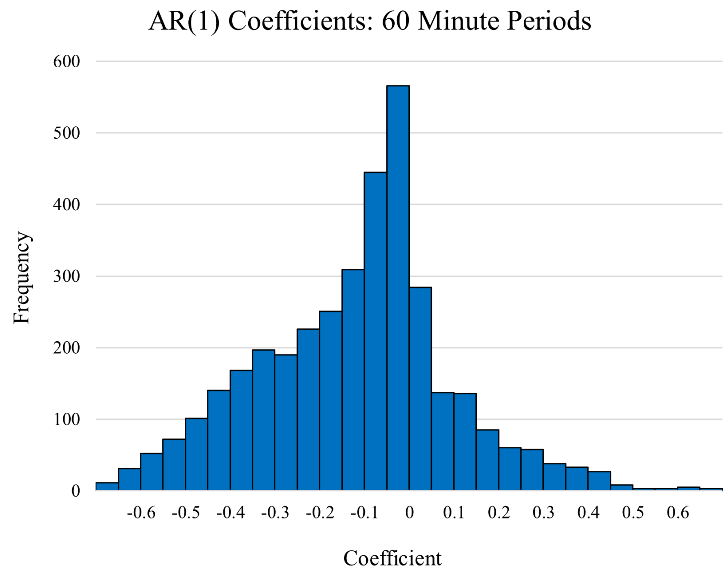

Figure 9 **A histogram of AR(1) coefficients for FanDuel's lines with 60 minute periods**

where σ_t^2 denotes the variance of differences between pairs of values that are t periods apart. The null hypothesis of the test is $VR(q) = 1$ for a given q. The test statistic $Z(q)$ is then given by:

$$Z(q) = \frac{VR(q) - 1}{\sqrt{\frac{2(2q-1)(q-1)}{3nq^2}}} \sim N(0,1), \tag{12}$$

where n is the number of periods in the time series. This test statistic is asymptotically $N(0,1)$, which is a useful property when testing only a single (and long) time series. In this case, because independent realizations of $Z(q)$ will be generated for every game, the sample mean is approximately normal, and we need only test whether it is equal to zero[9] for each value of q.

The variance ratios and test statistics were calculated with full line histories for all values of q from two to six for each sportsbook. The results are shown in Table 8. The mean variance ratio for each value of q was well below one for all four sportsbooks, again indicating negative autocorrelation. The mean ratios were decreasing in q; longer lags revealed stronger autocorrelations. As previously, if the games are treated as independent observations, the mean test statistic is clearly significant for every value of q.

Regardless of which of the preceding approaches is considered, the presence of negative autocorrelation is clear and substantial, thus revealing an inefficiency in these betting markets: when the line changes, it will (on average) move back somewhat toward its previous position. A likely explanation is that when a line change reduces its accuracy, sophisticated bettors will recognize

[9] Alternatively, it is possible to construct a corresponding test on the variance ratios directly to determine whether their means are different from one. Doing so yields similarly significant results.

	Lag (q)	Mean Variance Ratio	Mean $Z(q)$	p-value
Caesars	2	0.995	-0.073	0.013
	3	0.987	-0.159	<0.001
	4	0.978	-0.253	<0.001
	5	0.969	-0.341	<0.001
	6	0.960	-0.433	<0.001
DraftKings	2	0.978	-0.350	<0.001
	3	0.963	-0.499	<0.001
	4	0.947	-0.658	<0.001
	5	0.934	-0.793	<0.001
	6	0.921	-0.921	<0.001
FanDuel	2	0.936	-1.013	<0.001
	3	0.892	-1.396	<0.001
	4	0.858	-1.677	<0.001
	5	0.831	-1.913	<0.001
	6	0.807	-2.113	<0.001
PointsBet	2	0.968	-0.536	<0.001
	3	0.943	-0.786	<0.001
	4	0.922	-0.979	<0.001
	5	0.903	-1.170	<0.001
	6	0.884	-1.351	<0.001

Table 8 Variance ratio test results for all four sportsbooks.

this and take the other side of the bet, which prompts the sportsbook to move the line back in the other direction (FanDuel 2023). Note that the tests used in this section make no assertions about the magnitudes of subsequent changes, only that they are statistically significant. In particular, they do not imply the existence of a profitable simple betting strategy that involves betting on a team whose price has just decreased. They reflect a specific inefficiency that could be incorporated into more sophisticated strategies.

6. Discussion and Conclusion

This paper presented an analysis of detailed betting line movement from four sportsbooks for three MLB seasons, adding to a rich literature on sports betting markets that has relied predominantly on opening and closing lines only. At any given lead time, the market probability forecasts were quite reliable when the teams were evenly matched, with some small deviations for more lopsided games. Some simple profitable betting strategies arose involving lopsided games. The quality of forecasts declined significantly in the final 90 minutes before weekend day games, which is contrary to how we expect an efficient forecasting system to behave. The sequences of market probabilities within individual games clearly exhibited negatively autocorrelated changes. This result constitutes a violation of weak form market efficiency; that is, past prices contain information that is not fully reflected in the current price, and could be used to improve forecasts.

It is natural to ask how such pronounced negative autocorrelation can continue to exist in these markets. The most likely answer is that it does not enable any clearly profitable simple betting strategies. While negative autocorrelation presents lucrative opportunities in financial markets, sports bets cannot easily be sold in the way that most financial assets can; they are contracts between the bettor and the bookmaker. If a bettor knows that a (+130, -150) moneyline is likely to move slightly in a particular direction, the bettor cannot simply buy one side of the bet and then sell it for a profit after the line moves. In addition, the magnitude of the line movement is often not large enough to overcome fees on its own. The knowledge of nonzero expected future line movement would have to be incorporated into a more sophisticated strategy; that is, if the sophisticated strategy suggested betting on a team whose price had increased recently, it might be preferable instead to wait for the price to decrease again. An exploration of the kinds of strategies used by professional sports bettors is beyond the scope of this paper, but Miller and Davidow (2019) provide an excellent overview. It is certainly helpful for any bettor to be aware of a market inefficiency, but it is unlikely that a casual bettor could develop a profitable strategy based on this inefficiency alone.

Another possible explanation for the observed negative autocorrelation is that betting line changes might be beneficial to bookmakers. Line changes often prompt some additional bets (FanDuel 2023). As pointed out by Levitt (2004), Paul and Weinbach (2007), Paul and Weinbach (2008), and Kain and Logan (2014), bookmakers are trying to maximize profit, which does not always coincide perfectly with setting the most accurate line possible. A thorough exploration of this issue would require data on betting volume, but it raises the question of why market efficiency matters in this setting, as there is no clear axiomatic requirement for it. The most obvious reason to study it is that inefficiencies represent opportunities for bettors to develop strategies to exploit these markets. Another is that sportsbooks might want to reduce, eliminate, or otherwise address inefficiencies of which they are unaware. Market inefficiencies in this setting are also a point of broader interest; popular sports betting lines convey extremely accurate probability forecasts, and there is value in understanding what the shortcomings of those forecasts are and why they occur.

It is important to recognize that while betting lines at large sportsbooks are informed by models, they are set by human decision makers (FanDuel 2023). While these decision makers generally strive for their lines to be as accurate as possible (and are very successful, for the most part), they cannot be completely immune to biases. For example, *base-rate neglect* (Kahneman and Tversky 1972, Bar-Hillel 1980) could lead to over-weighting the impact of an injury announcement or a surprising lineup on a game's outcome, and *action bias* (Patt and Zeckhauser 2000, Bar-Eli et al. 2007) might increase the chances of a human moving a line when presented with information that does not warrant it. There are many different approaches for mitigating bias depending on the type and source (Fischoff 1981, Arkes 1991), but training and empirical feedback on performance are likely to be helpful in this setting.

It is also not guaranteed that employees' evaluations of tradeoffs and risk will always align perfectly with those of the sportsbook. For instance, large sportsbooks accept bets in an enormous number of markets, which allows them to be risk neutral with regard to most individual betting lines. If a sportsbook wishes its employees to act in a risk neutral manner when determining lines, it is crucial that the employees are evaluated and compensated in a way that disincentivizes risk averse behavior (e.g., setting lines with a primary goal of balancing volume).

In addition, most large sportsbooks determine their lines based at least partially on the lines of other large sportsbooks. This can amplify inefficiencies; when one sportsbook "incorrectly" moves a line, others may move their lines as well despite having received no other information that would have prompted them to do so. It is also possible that overreacting to noise is less costly than being too slow to react to a real signal. If so, a sportsbook would prefer that any mechanism for setting lines, whether human or algorithmic, err on the side of overreaction. Again, this is dependent on the behavior of the betting population, but would be a valuable direction for future work given access to betting volume data.

The question of whether simple profitable betting strategies exist arises frequently in analyses of sports betting, and was explored a few times in this paper with mixed results. Note that finding profitable betting strategies is a much more elusive goal than detecting market inefficiencies, as it must also overcome the bookmaker's fees. The primary challenge in finding such strategies is data limitations. As betting strategies are parameterized in increasing detail, they apply to fewer and fewer games, which makes it difficult to determine conclusively that they are indeed profitable. Even if one were to obtain many seasons of data, it is not clear that a very specific strategy that was profitable ten years ago would remain so today, given the evolution of these markets and the ability of sportsbooks to change their processes to address identified vulnerabilities. Section 2 provided a few examples of seemingly conflicting prior results that were based on data from different time periods.

Section 2 also included a brief comparison of sports betting markets and prediction markets. It should be noted that some prediction markets for major sporting events do exist. For instance, betfair.com provides exchange markets for a wide variety of events, including sports. If detailed price movement data from those markets could be obtained, it would be informative to discover which of the results in this paper hold in prediction markets for the same (or comparable) events.

Another potential direction for future work is to explore how in-game betting lines behave, and whether the findings in this paper arise in those markets as well. In-game markets differ from pre-game betting markets in several ways; the speed and magnitude of information revelation are much higher, and there is a non-trivial difference in latency regarding how quickly information is received by the various participants.

It would also be useful to explore how widely this paper's findings can be generalized to other sports or other bet types. Negative autocorrelation, in particular, does not require the betting market to be either MLB or moneylines, and might arise in a variety of settings. That would be a valuable avenue of work for researchers who are able to obtain comparable data on betting line movement. Currently, the primary hurdle in doing this type of analysis is the lack of access to sufficient detailed data, but given the enormous recent growth in the sports betting industry, it is likely that such data will become increasingly available in the coming years.

Acknowledgements

I am grateful to members of FanDuel's Risk & Trading department for sharing their experiences and insights, and to an Associate Editor and two reviewers for their valuable comments.

References

Andrikogiannopoulou, A. and Papakonstantinou, F. (2018). Individual reaction to past performance sequences: evidence from a real marketplace. *Management Science*, 64(4):1957–1973.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, 110(3):486.

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., et al. (2008). The promise of prediction markets. *Science*, 320(5878):877–878.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., Ungar, L., and Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science*, 63(3):691–706.

Bar-Eli, M., Azar, O. H., Ritov, I., Keidar-Levin, Y., and Schein, G. (2007). Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of Economic Psychology*, 28(5):606–621.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3):211–233.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Bennett, R. W. (2021). Next game reaction to mispriced betting lines in college football. *Applied Economics Letters*, 28(12):1006–1009.

Berkowitz, J. P., Depken, C. A., and Gandar, J. M. (2018). The conversion of money lines into win probabilities: Reconciliations and simplifications. *Journal of Sports Economics*, 19(7):990–1015.

Bickel, J. E. and Kim, S. D. (2014). Re-examining the efficiency of the major league baseball over-under betting market. *Applied Financial Economics*, 24(18):1229–1234.

Bowman, R. A., Ashman, T., and Lambrinos, J. (2013). Prospective measures of competitive balance: Application to money lines in Major League Baseball. *Applied Economics*, 45(29):4071–4081.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3.

Brown, K. H. and Abraham, F. J. (2002). Testing market efficiency in the major league baseball over-under betting market. *Journal of Sports Economics*, 3(4):311–319.

Can, E. and Nichols, M. W. (2022). The income elasticity of gross sports betting revenues in Nevada: Short-run and long-run estimates. *Journal of Sports Economics*, 23(2):175–199.

Charles, A. and Darné, O. (2009). Variance-ratio tests of random walk: an overview. *Journal of Economic Surveys*, 23(3):503–527.

Conrad, J. and Kaul, G. (1988). Time-variation in expected returns. *Journal of Business*, pages 409–425.

Cutler, D. M., Poterba, J. M., and Summers, L. H. (1991). Speculative dynamics. *The Review of Economic Studies*, 58(3):529–546.

Dana, J. D. and Knetter, M. M. (1994). Learning and efficiency in a gambling market. *Management Science*, 40(10):1317–1328.

Dare, W. H., Dennis, S. A., and Paul, R. J. (2015). Player absence and betting lines in the NBA. *Finance Research Letters*, 13:130–136.

De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3):793–805.

De Finetti, B. (1962). Does it make sense to speak of 'good probability appraisers'. *The scientist speculates: An anthology of partly-baked ideas*, pages 257–364.

Drèze, X. and Nunes, J. C. (2004). Using combined-currency prices to lower consumers' perceived cost. *Journal of Marketing Research*, 41(1):59–72.

Durand, R. B., Patterson, F. M., and Shank, C. A. (2021). Behavioral biases in the nfl gambling market: Overreaction to news and the recency bias. *Journal of Behavioral and Experimental Finance*, 31:100522.

Estelami, H. (2003). The effect of price presentation tactics on consumer evaluation effort of multi-dimensional prices. *Journal of Marketing Theory and Practice*, 11(2):1–16.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. and French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96(2):246–273.

FanDuel (2023). Risk & Trading department. Personal communication.

Fischoff, B. (1981). Debiasing. Technical report, DECISION RESEARCH EUGENE OR.

Friedman, D. (1983). Effective scoring rules for probabilistic forecasts. *Management Science*, 29(4):447–454.

Gandar, J. M., Dare, W. H., Brown, C. R., and Zuber, R. A. (1998). Informed traders and price variations in the betting market for professional basketball games. *The Journal of Finance*, 53(1):385–401.

Gandar, J. M., Zuber, R. A., Johnson, R. S., and Dare, W. (2002). Re-examining the betting market on major league baseball games: is there a reverse favourite-longshot bias? *Applied Economics*, 34(10):1309–1317.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Grand View Research (2021). Sports Betting Market Size, Share & Trends Analysis Report By Platform (Online, Offline), By Type (Fixed Odds Wagering, eSports Betting), By Sports Type (Football, Basketball), By Region, And Segment Forecasts, 2021 - 2028. Accessed 1/22/2022, https://www.grandviewresearch.com/industry-analysis/sports-betting-market-report.

Green, E., Lee, H., and Rothschild, D. M. (2020). The favorite-longshot midas. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Huang, B.-N. (1995). Do Asian stock market prices follow random walks? Evidence from the variance ratio test. *Applied Financial Economics*, 5(4):251–256.

Jose, V. R. R., Nau, R. F., and Winkler, R. L. (2008). Scoring rules, generalized entropy, and utility maximization. *Operations Research*, 56(5):1146–1157.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3):430–454.

Kain, K. J. and Logan, T. D. (2014). Are sports betting markets prediction markets? Evidence from a new test. *Journal of Sports Economics*, 15(1):45–63.

Lahiri, K. and Wang, J. G. (2013). Evaluating probability forecasts for gdp declines using alternative methodologies. *International Journal of Forecasting*, 29(1):175–190.

Larrick, R. P. and Soll, J. B. (2008). The MPG illusion. *Science*, 320(5883):1593–1594.

Larson, R. B. (2014). Psychological pricing principles for organizations with market power. *Journal of Applied Business and Economics*, 16(1):11–25.

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets? *The Economic Journal*, 114(495):223–246.

Lichtenstein, S. and Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, 26(2):149–171.

Liu, C. Y. and He, J. (1991). A variance-ratio test of random walks in foreign exchange rates. *The Journal of Finance*, 46(2):773–785.

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1):41–66.

Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters*, 91(3):425–429.

Miller, E. and Davidow, M. (2019). *The Logic Of Sports Betting*.

Miller Jr, T. W. and Rapach, D. E. (2013). An intra-week efficiency analysis of bookie-quoted NFL betting lines in NYC. *Journal of Empirical Finance*, 24:10–23.

Morewedge, C. K., Tang, S., and Larrick, R. P. (2018). Betting your favorite to win: Costly reluctance to hedge desired outcomes. *Management Science*, 64(3):997–1014.

Moskowitz, T. J. (2021). Asset pricing and sports betting. *The Journal of Finance*, 76(6):3153–3209.

Murphy, A. H. and Winkler, R. L. (1970). Scoring rules in probability assessment and evaluation. *Acta Psychologica*, 34:273–286.

Nordhaus, W. D. (1987). Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics*, pages 667–674.

Oxford Analytica (2022). US states see sports betting as new revenue source. *Emerald Expert Briefings*.

Patt, A. and Zeckhauser, R. (2000). Action bias and environmental decisions. *Journal of Risk and Uncertainty*, 21:45–72.

Paul, R., Weinbach, A. P., Borghesi, R., and Wilson, M. (2008). Using betting market odds to measure the uncertainty of outcome in major league baseball. *International Journal of Sport Finance*, 4(4):255–263.

Paul, R. J. and Weinbach, A. P. (2004). Testing market efficiency in the major league baseball over-under betting market. *Journal of Sports Economics*, 5(1):93–95.

Paul, R. J. and Weinbach, A. P. (2007). Does sportsbook. com set pointspreads to maximize profits? *The Journal of Prediction Markets*, 1(3):209–218.

Paul, R. J. and Weinbach, A. P. (2008). Price setting in the NBA gambling market: Tests of the Levitt model of sportsbook behavior. *International Journal of Sport Finance*, 3(3):137–145.

Peine, K., Heitmann, M., and Herrmann, A. (2009). Getting a feel for price affect: a conceptual framework and empirical investigation of consumers' emotional responses to price information. *Psychology & Marketing*, 26(1):39–66.

Poterba, J. M. and Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22(1):27–59.

Prelec, D. (1998). The probability weighting function. *Econometrica*, pages 497–527.

Rascher, D. A. (1999). A test of the optimal positive production network externality in Major League Baseball. *Sports, Economics: Current Research.*

Regnier, E. (2018). Probability forecasts made at multiple lead times. *Management Science*, 64(5):2407–2426.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Management Review*, 6(2).

Sauer, R. D. (2005). The state of research on markets for sports betting and suggested future directions. *Journal of Economics and Finance*, 29(3):416–426.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801.

Schnytzer, A. and Weinberg, G. (2008). Testing for home team and favorite biases in the australian rules football fixed-odds and point spread betting markets. *Journal of Sports Economics*, 9(2):173–190.

Seillier-Moiseiwitsch, F. and Dawid, A. (1993). On testing the validity of sequential probability forecasts. *Journal of the American Statistical Association*, 88(421):355–359.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1(1):43–61.

Seybert, N. and Bloomfield, R. (2009). Contagion of wishful thinking in markets. *Management Science*, 55(5):738–751.

Shank, C. A. (2018). Is the NFL betting market still inefficient? *Journal of Economics and Finance*, 42(4):818–827.

Sinkey, M. and Logan, T. (2014). Does the hot hand drive the market? evidence from college football betting markets. *Eastern Economic Journal*, 40(4):583–603.

Smith, G. and Ryoo, H.-J. (2003). Variance ratio tests of the random walk hypothesis for european emerging stock markets. *The European Journal of Finance*, 9(3):290–300.

Summers, L. H. (1986). Does the stock market rationally reflect fundamental values? *The Journal of Finance*, 41(3):591–601.

Swartz, M. (2009). Home-field advantage. *The Baseball Research Journal*, 38(2):31–47.

Vergin, R. C. and Sosik, J. J. (1999). No place like home: an examination of the home field advantage in gambling strategies in NFL football. *Journal of Economics and Business*, 51(1):21–31.

Wever, S. and Aadland, D. (2012). Herd behaviour and underdogs in the NFL. *Applied Economics Letters*, 19(1):93–97.

Winkler, R. L. (1967). The quantification of judgment: Some methodological suggestions. *Journal of the American Statistical Association*, 62(320):1105–1120.

Wolfers, J. and Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126.

Woodland, L. M. and Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias: the baseball betting market. *The Journal of Finance*, 49(1):269–279.

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management science*, 42(12):1676–1690.

Zion Market Research (2021). Sports Betting Market - By Platform (Online And Offline), By Type (Line-In-Play, Fixed Old Betting, Exchange Betting, Daily Fantasy, Spread Betting, E-Sports, Pari-Mutuel, And Others), And By Sports Type (Football, Baseball, Basketball, Hockey, Cricket, Tennis, Golf, Boxing, Horse Riding, Auto Racing, And Others): Global Industry Perspective, Comprehensive Analysis, And Forecast, 2021-2028. Accessed 1/22/2022, https://www.zionmarketresearch.com/report/sports-betting-market.

Appendix A. Reliability Results

Reliability charts for Caesars, DraftKings, and FanDuel are included here.

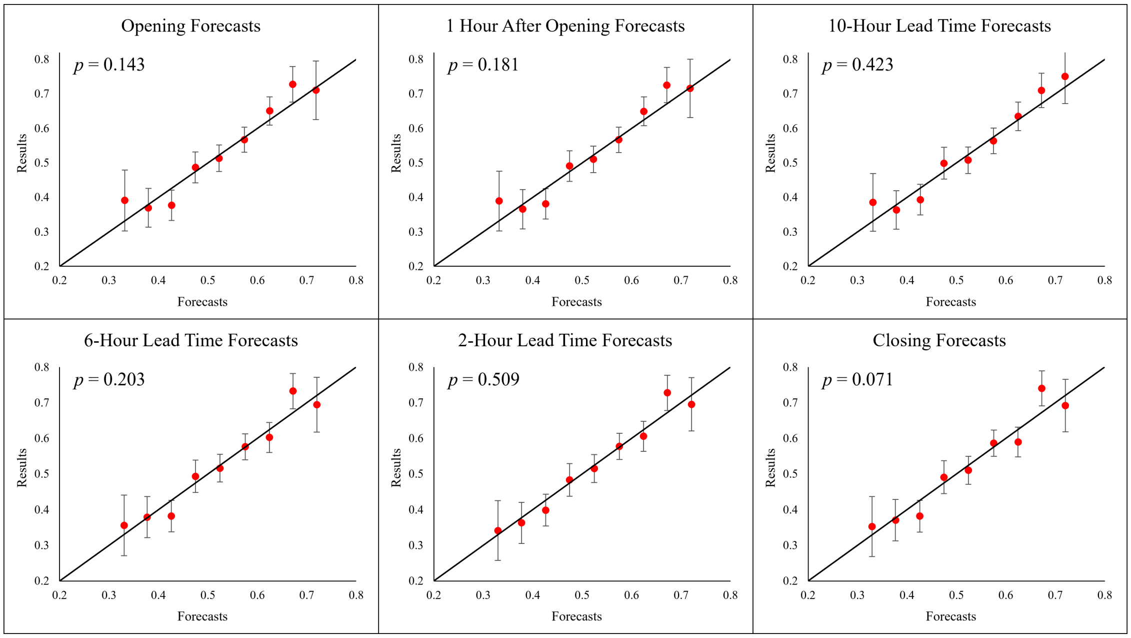

Figure 10 **Reliability of Caesars' implied home team win probability forecasts at six different times.**

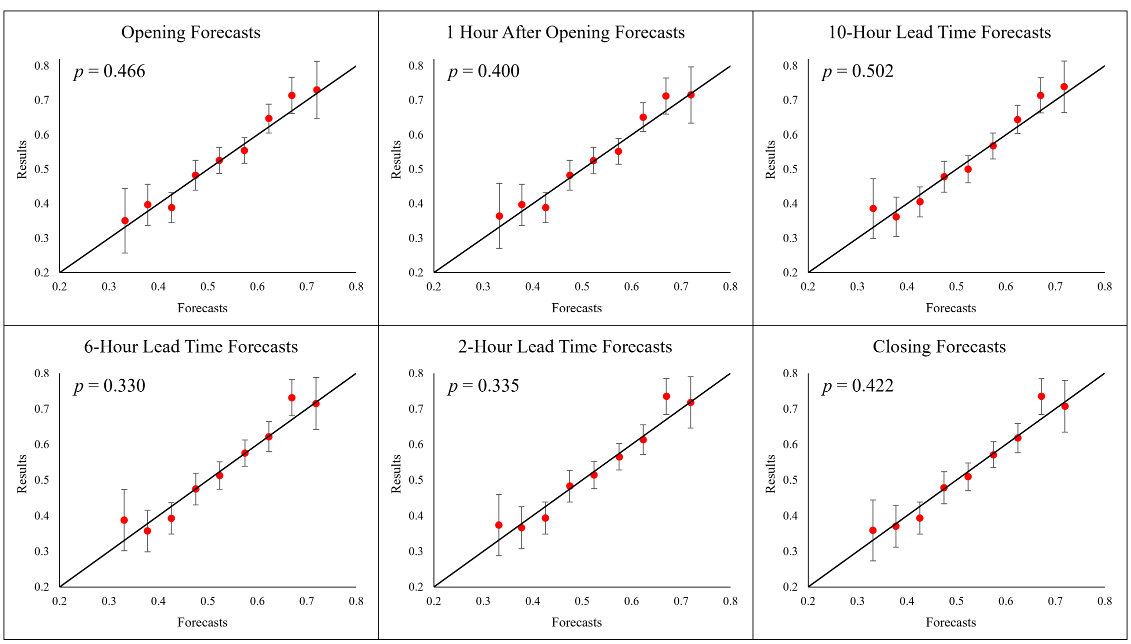

Figure 11 **Reliability of DraftKings' implied home team win probability forecasts at six different times.**

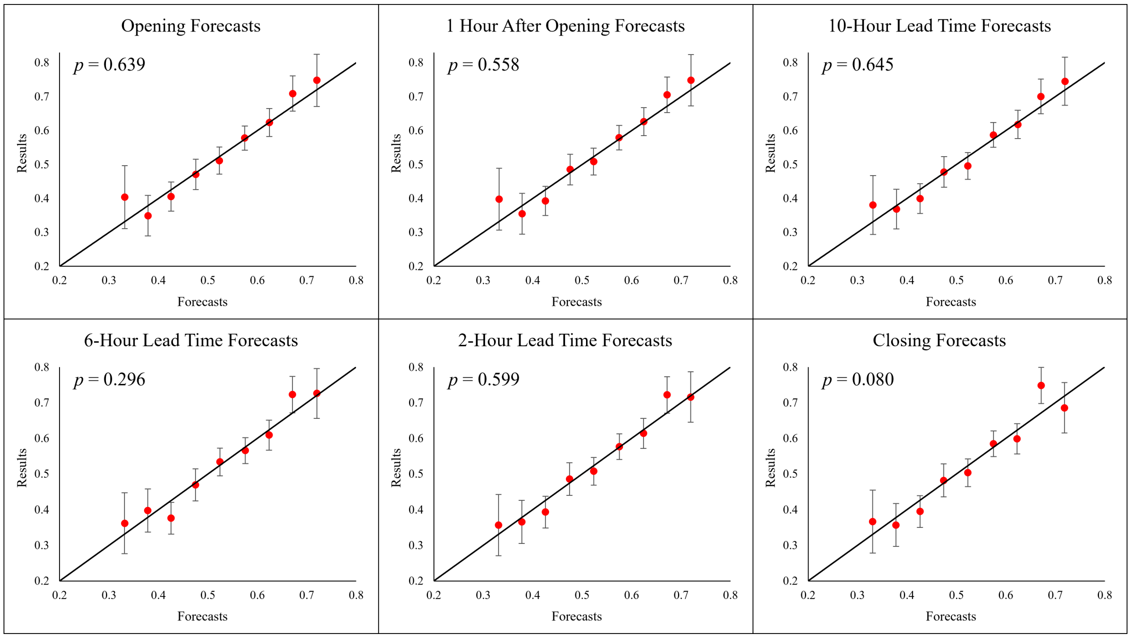

Figure 12 Reliability of FanDuel's implied home team win probability forecasts at six different times.

Appendix B. Simple Betting Strategies

The results for the 360 simple betting strategies for Caesars, DraftKings, and FanDuel are included here.

	Bet on home team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
[+340, +145)	0.032 (0.189)	0.025 (0.219)	0.03 (0.19)	0.036 (0.168)	0.014 (0.265)	0.023 (0.224)
[+145, +120)	-0.131 (0.962)	-0.117 (0.941)	-0.117 (0.936)	-0.149 (0.979)	-0.141 (0.971)	-0.142 (0.974)
[+120, +104)	-0.04 (0.583)	-0.04 (0.582)	-0.068 (0.765)	-0.081 (0.834)	-0.101 (0.909)	-0.102 (0.907)
[+104, −110)	-0.023 (0.46)	-0.01 (0.368)	0.014 (0.221)	0.026 (0.168)	**0.105** **(0.011)**	**0.072** **(0.041)**
[−110, −125)	-0.039 (0.569)	-0.049 (0.651)	-0.071 (0.803)	-0.075 (0.827)	-0.102 (0.934)	-0.076 (0.838)
[−125, −140)	0.008 (0.206)	0.002 (0.248)	0.042 (0.058)	0.028 (0.1)	0.008 (0.214)	-0.011 (0.361)
[−140, −160)	-0.096 (0.953)	-0.087 (0.929)	-0.085 (0.925)	-0.032 (0.577)	-0.004 (0.305)	0.024 (0.118)
[−160, −180)	-0.025 (0.491)	-0.036 (0.595)	-0.055 (0.756)	-0.087 (0.926)	-0.118 (0.986)	-0.128 (0.993)
[−180, −220)	**0.059** **(0.013)**	**0.065** **(0.008)**	0.032 (0.063)	**0.04** **(0.041)**	**0.039** **(0.044)**	0.027 (0.085)
[−220, −600]	-0.015 (0.304)	-0.012 (0.275)	-0.003 (0.187)	-0.017 (0.313)	-0.001 (0.166)	-0.005 (0.192)
	Bet on visiting team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
[+435, +185)	-0.098 (0.794)	-0.106 (0.822)	-0.124 (0.885)	-0.095 (0.801)	-0.116 (0.872)	-0.109 (0.855)
[+185, +155)	-0.151 (0.977)	-0.149 (0.975)	-0.114 (0.916)	-0.096 (0.86)	-0.057 (0.683)	-0.065 (0.726)
[+155, +138)	0.031 (0.196)	0.053 (0.113)	0.075 (0.054)	0.07 (0.064)	**0.079** **(0.043)**	**0.094** **(0.025)**
[+138, +120)	-0.018 (0.46)	-0.042 (0.622)	-0.002 (0.361)	-0.026 (0.517)	-0.036 (0.581)	-0.056 (0.705)
[+120, +110)	-0.079 (0.785)	-0.058 (0.679)	-0.13 (0.939)	-0.113 (0.9)	-0.115 (0.907)	-0.112 (0.9)
[+110, −105)	-0.028 (0.489)	-0.03 (0.503)	0.01 (0.216)	-0.015 (0.383)	-0.02 (0.425)	0.016 (0.188)
[−105, −120)	0.002 (0.245)	0.007 (0.215)	-0.067 (0.764)	-0.036 (0.548)	-0.063 (0.736)	-0.083 (0.846)
[−120, −140)	-0.033 (0.54)	-0.03 (0.512)	-0.005 (0.298)	0.013 (0.173)	0.036 (0.075)	0.036 (0.081)
[−140, −168)	**0.048** **(0.039)**	0.041 (0.057)	**0.046** **(0.048)**	**0.046** **(0.047)**	0.036 (0.079)	**0.045** **(0.049)**
[−168, −424]	-0.047 (0.678)	-0.052 (0.722)	-0.046 (0.674)	-0.066 (0.834)	-0.052 (0.727)	-0.06 (0.79)

Table 9 **Returns and p-values (in parentheses) for 120 simple betting strategies on Caesars, results with p-values of 0.05 or lower bolded**

	Bet on home team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
$[+340, +145)$	0.01 (0.265)	0.019 (0.226)	-0.002 (0.315)	0.049 (0.115)	0.012 (0.249)	0.027 (0.187)
$[+145, +120)$	-0.104 (0.877)	-0.104 (0.877)	-0.094 (0.845)	-0.133 (0.948)	-0.117 (0.913)	-0.159 (0.981)
$[+120, +104)$	-0.067 (0.715)	-0.063 (0.687)	-0.056 (0.635)	-0.112 (0.915)	-0.122 (0.94)	-0.106 (0.894)
$[+104, -110)$	-0.045 (0.539)	-0.05 (0.576)	-0.033 (0.456)	-0.007 (0.283)	0.051 (0.058)	0.047 (0.07)
$[-110, -125)$	-0.008 (0.237)	-0.012 (0.264)	-0.071 (0.733)	-0.101 (0.894)	-0.115 (0.939)	-0.067 (0.714)
$[-125, -140)$	-0.062 (0.713)	-0.062 (0.713)	-0.034 (0.473)	0.021 (0.09)	-0.015 (0.307)	-0.021 (0.359)
$[-140, -160)$	-0.04 (0.555)	-0.044 (0.593)	-0.036 (0.514)	-0.032 (0.473)	-0.014 (0.299)	-0.022 (0.372)
$[-160, -180)$	-0.039 (0.54)	-0.029 (0.453)	-0.097 (0.913)	-0.107 (0.943)	-0.085 (0.863)	-0.113 (0.956)
$[-180, -220)$	**0.024 (0.049)**	**0.028 (0.038)**	0.005 (0.124)	0.005 (0.128)	-0.015 (0.278)	-0.01 (0.236)
$[-220, -600]$	-0.015 (0.242)	-0.024 (0.333)	**0.023 (0.029)**	0.003 (0.095)	0.007 (0.072)	0.003 (0.089)

	Bet on visiting team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
$[+435, +185)$	-0.037 (0.494)	-0.01 (0.37)	-0.095 (0.76)	-0.055 (0.582)	-0.091 (0.749)	-0.093 (0.758)
$[+185, +155)$	-0.116 (0.87)	-0.135 (0.919)	-0.196 (0.991)	-0.195 (0.99)	-0.181 (0.982)	-0.152 (0.953)
$[+155, +138)$	-0.113 (0.899)	-0.121 (0.919)	0.045 (0.1)	0.04 (0.117)	**0.073 (0.042)**	**0.097 (0.018)**
$[+138, +120)$	0.023 (0.17)	0.038 (0.114)	0.016 (0.202)	-0.017 (0.386)	-0.075 (0.757)	-0.056 (0.647)
$[+120, +110)$	-0.025 (0.434)	-0.037 (0.508)	-0.113 (0.887)	-0.113 (0.887)	-0.009 (0.341)	-0.145 (0.951)
$[+110, -105)$	-0.09 (0.845)	-0.085 (0.817)	0.003 (0.205)	-0.004 (0.244)	-0.043 (0.537)	0.003 (0.203)
$[-105, -120)$	-0.024 (0.36)	-0.026 (0.376)	-0.051 (0.583)	-0.068 (0.71)	-0.039 (0.488)	-0.083 (0.797)
$[-120, -140)$	-0.027 (0.392)	-0.016 (0.296)	-0.037 (0.487)	0.015 (0.113)	-0.013 (0.282)	0.007 (0.152)
$[-140, -168)$	**0.033 (0.05)**	0.022 (0.083)	0.007 (0.153)	0.026 (0.069)	0.033 (0.051)	**0.049 (0.021)**
$[-168, -424]$	-0.075 (0.844)	-0.077 (0.855)	-0.043 (0.573)	-0.071 (0.819)	-0.051 (0.652)	-0.058 (0.718)

Table 10 **Returns and p-values (in parentheses) for 120 simple betting strategies on DraftKings, results with p-values of 0.05 or lower bolded**

	Bet on home team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
$[+340, +145)$	0.007 (0.266)	-0.015 (0.374)	0.018 (0.213)	0.016 (0.224)	0.005 (0.272)	-0.006 (0.326)
$[+145, +120)$	-0.107 (0.877)	-0.115 (0.901)	-0.128 (0.928)	-0.126 (0.927)	-0.13 (0.934)	-0.116 (0.898)
$[+120, +104)$	-0.054 (0.614)	-0.025 (0.412)	-0.079 (0.774)	-0.105 (0.884)	-0.128 (0.944)	-0.124 (0.938)
$[+104, -110)$	-0.07 (0.744)	-0.04 (0.52)	-0.002 (0.242)	0.015 (0.157)	0.026 (0.113)	**0.059** **(0.034)**
$[-110, -125)$	-0.07 (0.737)	-0.133 (0.969)	-0.111 (0.929)	-0.109 (0.926)	-0.111 (0.934)	-0.107 (0.922)
$[-125, -140)$	0.027 (0.077)	0.035 (0.059)	0.01 (0.156)	**0.043** **(0.038)**	**0.051** **(0.027)**	-0.044 (0.561)
$[-140, -160)$	-0.091 (0.901)	-0.083 (0.867)	-0.045 (0.571)	-0.072 (0.794)	-0.037 (0.492)	0.01 (0.123)
$[-160, -180)$	-0.031 (0.429)	-0.029 (0.412)	-0.087 (0.867)	-0.073 (0.799)	-0.104 (0.94)	-0.11 (0.958)
$[-180, -220)$	-0.007 (0.194)	-0.021 (0.317)	-0.013 (0.241)	-0.024 (0.355)	-0.021 (0.325)	-0.019 (0.309)
$[-220, -600]$	-0.008 (0.176)	0.009 (0.073)	0.004 (0.089)	-0.005 (0.14)	-0.003 (0.124)	0 (0.104)
	Bet on visiting team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
$[+435, +185)$	-0.087 (0.731)	-0.091 (0.749)	-0.109 (0.823)	-0.101 (0.8)	-0.084 (0.728)	-0.111 (0.837)
$[+185, +155)$	-0.069 (0.678)	-0.091 (0.784)	-0.135 (0.926)	-0.098 (0.809)	-0.123 (0.893)	-0.141 (0.927)
$[+155, +138)$	-0.101 (0.822)	-0.06 (0.625)	0.052 (0.096)	0.03 (0.16)	0.058 (0.081)	**0.136** **(0.005)**
$[+138, +120)$	0.041 (0.071)	0.032 (0.097)	-0.012 (0.316)	-0.015 (0.334)	-0.033 (0.463)	-0.083 (0.8)
$[+120, +110)$	-0.132 (0.931)	-0.133 (0.936)	-0.109 (0.875)	-0.044 (0.543)	-0.05 (0.58)	-0.056 (0.621)
$[+110, -105)$	-0.04 (0.52)	-0.01 (0.283)	0.025 (0.098)	-0.022 (0.372)	-0.019 (0.344)	0.026 (0.091)
$[-105, -120)$	0.011 (0.173)	-0.039 (0.513)	-0.079 (0.788)	-0.119 (0.941)	-0.083 (0.801)	-0.129 (0.955)
$[-120, -140)$	-0.06 (0.689)	-0.051 (0.618)	0.006 (0.171)	**0.044** **(0.039)**	-0.003 (0.227)	0.036 (0.052)
$[-140, -168)$	0.027 (0.056)	**0.037** **(0.034)**	0.019 (0.092)	**0.038** **(0.036)**	0.019 (0.09)	-0.005 (0.222)
$[-168, -424]$	-0.046 (0.577)	-0.046 (0.578)	-0.061 (0.73)	-0.084 (0.889)	-0.044 (0.56)	-0.037 (0.484)

Table 11 **Returns and p-values (in parentheses) for 120 simple betting strategies on FanDuel, results with p-values of 0.05 or lower bolded**

Appendix C. Split-Sample Betting Strategy Results

The results for the 120 simple betting strategies on PointsBet are included here for the set of games split randomly into two equal-sized subsets.

Moneyline Range:	Bet on home team at lead time (hours): Open	Open+1	10	6	2	0
$[+340, +145)$	0.022 (0.266)	-0.008 (0.368)	0.008 (0.302)	0.043 (0.191)	0.06 (0.149)	0.059 (0.15)
$[+145, +120)$	-0.098 (0.776)	-0.085 (0.721)	-0.087 (0.726)	-0.125 (0.853)	-0.2 (0.977)	-0.229 (0.991)
$[+120, +104)$	-0.072 (0.666)	-0.088 (0.738)	-0.111 (0.819)	-0.097 (0.769)	-0.089 (0.733)	-0.071 (0.65)
$[+104, -110)$	-0.079 (0.649)	0.014 (0.258)	0.095 (0.06)	-0.005 (0.339)	0.067 (0.135)	0.063 (0.136)
$[-110, -125)$	-0.127 (0.893)	-0.137 (0.917)	-0.052 (0.543)	-0.096 (0.774)	-0.056 (0.566)	-0.047 (0.514)
$[-125, -140)$	**0.075** **(0.037)**	0.027 (0.148)	-0.109 (0.866)	0.002 (0.249)	0.022 (0.156)	0.012 (0.205)
$[-140, -160)$	-0.075 (0.74)	-0.037 (0.508)	0.033 (0.135)	0.012 (0.217)	-0.058 (0.641)	-0.049 (0.583)
$[-160, -180)$	-0.069 (0.69)	-0.075 (0.719)	-0.147 (0.958)	-0.169 (0.984)	-0.12 (0.907)	-0.128 (0.926)
$[-180, -220)$	0.008 (0.157)	-0.006 (0.232)	-0.002 (0.21)	-0.039 (0.483)	-0.079 (0.777)	-0.066 (0.694)
$[-220, -600]$	**0.043** **(0.027)**	**0.053** **(0.017)**	**0.033** **(0.037)**	**0.059** **(0.007)**	**0.068** **(0.004)**	**0.059** **(0.006)**

Moneyline Range:	Bet on visiting team at lead time (hours): Open	Open+1	10	6	2	0
$[+435, +185)$	-0.095 (0.655)	-0.112 (0.704)	-0.096 (0.662)	-0.131 (0.757)	-0.165 (0.843)	-0.166 (0.851)
$[+185, +155)$	-0.27 (0.992)	-0.275 (0.993)	-0.255 (0.991)	-0.206 (0.968)	-0.257 (0.991)	-0.225 (0.977)
$[+155, +138)$	-0.061 (0.596)	-0.079 (0.669)	0.014 (0.279)	0.058 (0.144)	**0.134** **(0.025)**	**0.113** **(0.041)**
$[+138, +120)$	**0.1** **(0.042)**	0.085 (0.063)	0.055 (0.144)	-0.027 (0.452)	0.018 (0.254)	-0.027 (0.452)
$[+120, +110)$	-0.265 (0.993)	-0.194 (0.959)	-0.168 (0.925)	0.012 (0.301)	-0.079 (0.666)	-0.008 (0.382)
$[+110, -105)$	0.016 (0.23)	0.034 (0.167)	0.031 (0.163)	-0.097 (0.78)	-0.079 (0.705)	-0.106 (0.819)
$[-105, -120)$	0.022 (0.154)	0.011 (0.2)	-0.02 (0.359)	-0.081 (0.688)	-0.098 (0.767)	-0.077 (0.671)
$[-120, -140)$	-0.13 (0.918)	-0.165 (0.974)	-0.098 (0.81)	0.005 (0.235)	0.025 (0.157)	0.039 (0.118)
$[-140, -168)$	0.056 (0.051)	**0.083** **(0.017)**	0.018 (0.166)	0.015 (0.183)	0.039 (0.099)	0 (0.263)
$[-168, -424]$	-0.045 (0.524)	-0.051 (0.57)	-0.051 (0.576)	-0.042 (0.502)	-0.061 (0.651)	-0.03 (0.406)

Table 12 **Returns and p-values (in parentheses) for 120 simple betting strategies on PointsBet using the first random subset of games, results with p-values of 0.05 or lower bolded**

	Bet on home team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
[+340, +145)	0.095 (0.105)	0.051 (0.197)	0.016 (0.288)	-0.012 (0.387)	-0.004 (0.357)	0.042 (0.21)
[+145, +120)	-0.146 (0.908)	-0.119 (0.835)	-0.104 (0.779)	-0.084 (0.701)	-0.098 (0.751)	-0.154 (0.905)
[+120, +104)	-0.053 (0.576)	-0.081 (0.721)	-0.08 (0.715)	-0.137 (0.914)	-0.139 (0.913)	-0.08 (0.714)
[+104, −110)	-0.059 (0.568)	-0.052 (0.536)	-0.027 (0.419)	0.034 (0.181)	0.094 (0.059)	0.044 (0.175)
[−110, −125)	-0.06 (0.588)	-0.021 (0.362)	-0.041 (0.479)	-0.11 (0.832)	-0.117 (0.86)	-0.143 (0.931)
[−125, −140)	-0.049 (0.561)	-0.049 (0.558)	-0.037 (0.482)	0 (0.259)	-0.047 (0.547)	-0.034 (0.464)
[−140, −160)	-0.009 (0.326)	-0.023 (0.414)	-0.041 (0.531)	0.015 (0.212)	0.012 (0.23)	0.022 (0.183)
[−160, −180)	-0.177 (0.986)	-0.144 (0.956)	-0.077 (0.728)	-0.049 (0.566)	-0.002 (0.267)	-0.025 (0.408)
[−180, −220)	**0.045** **(0.04)**	0.022 (0.097)	-0.022 (0.352)	-0.054 (0.604)	-0.06 (0.652)	-0.044 (0.528)
[−220, −600]	-0.02 (0.289)	-0.017 (0.27)	-0.021 (0.285)	-0.036 (0.41)	-0.054 (0.572)	-0.062 (0.646)
	Bet on visiting team at lead time (hours):					
Moneyline Range:	Open	Open+1	10	6	2	0
[+435, +185)	-0.091 (0.633)	-0.097 (0.65)	-0.038 (0.476)	0.077 (0.155)	0.04 (0.232)	0.062 (0.172)
[+185, +155)	-0.104 (0.748)	-0.075 (0.637)	-0.066 (0.602)	-0.139 (0.855)	-0.163 (0.903)	-0.153 (0.881)
[+155, +138)	-0.022 (0.422)	-0.054 (0.566)	-0.077 (0.656)	0 (0.324)	0.069 (0.103)	0.06 (0.125)
[+138, +120)	-0.041 (0.52)	-0.028 (0.458)	-0.04 (0.513)	-0.09 (0.731)	-0.146 (0.902)	-0.167 (0.937)
[+120, +110)	-0.005 (0.366)	-0.029 (0.465)	-0.064 (0.613)	-0.137 (0.845)	-0.116 (0.786)	-0.101 (0.752)
[+110, −105)	-0.024 (0.402)	-0.024 (0.403)	-0.014 (0.351)	-0.009 (0.321)	0.033 (0.15)	0.019 (0.206)
[−105, −120)	-0.082 (0.702)	-0.134 (0.9)	-0.076 (0.671)	-0.006 (0.289)	0 (0.262)	0.009 (0.226)
[−120, −140)	-0.009 (0.293)	0.038 (0.086)	-0.009 (0.293)	-0.06 (0.628)	-0.028 (0.412)	-0.022 (0.371)
[−140, −168)	0.044 (0.08)	0.045 (0.082)	0.03 (0.131)	0.059 (0.053)	-0.025 (0.421)	0.002 (0.262)
[−168, −424]	-0.109 (0.89)	-0.106 (0.884)	-0.063 (0.657)	-0.067 (0.68)	-0.023 (0.359)	-0.046 (0.529)

Table 13 **Returns and p-values (in parentheses) for 120 simple betting strategies on PointsBet using the second random subset of games, results with p-values of 0.05 or lower bolded**

Appendix D. Scoring Results

Mean quadratic scores for weeknight and weekend day games for Caesars, FanDuel, and PointsBet are included here. Note that these three sportsbooks' lines tend to open later than DraftKings' lines, which leads to the higher performance during the first three hours after opening observed in these charts.

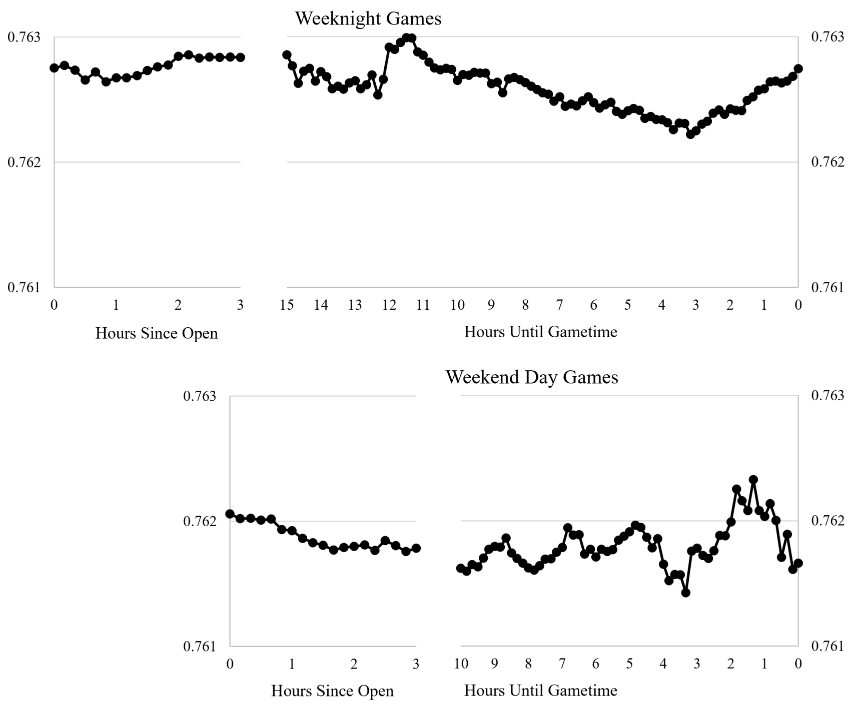

Figure 13 **Performance of Caesars' lines for weeknight games and weekend day games.**

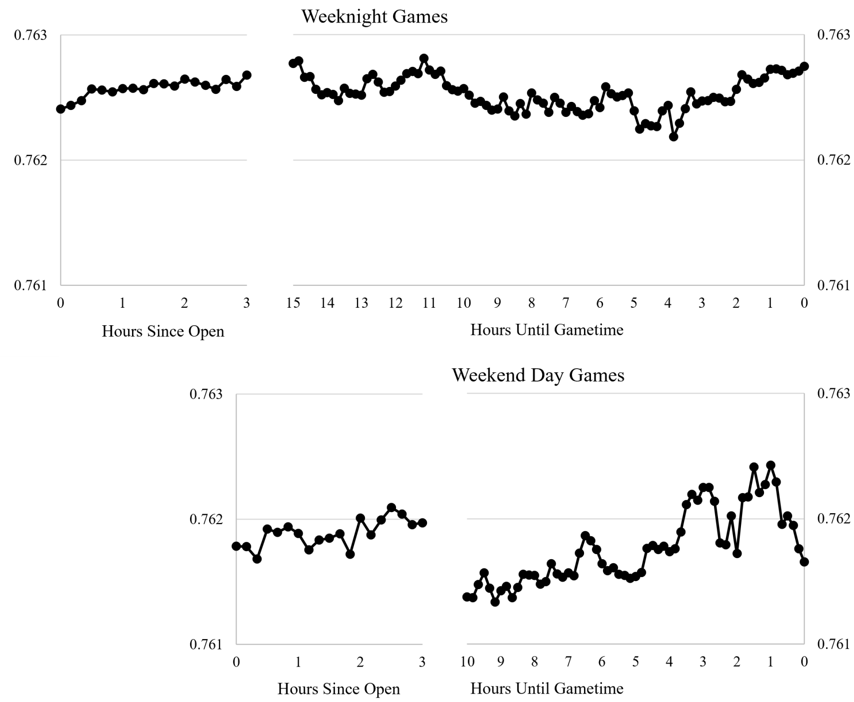

Figure 14 **Performance of FanDuel's lines for weeknight games and weekend day games.**

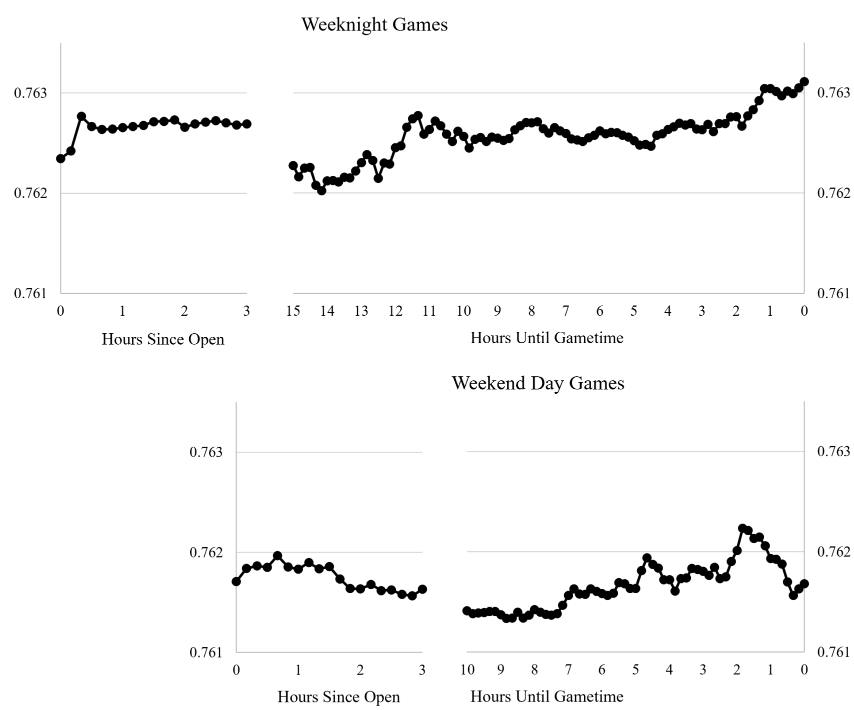

Figure 15 **Performance of PointsBet' lines for weeknight games and weekend day games.**