Economics Bulletin

Volume 42, Issue 3

Is herding efficient? Evidence from the college football point spread market

Alexander Traugutt
Elon University

Jarid Morton
Western Kentucky University

Abstract

This study examined the efficiency of herd behavior in the college football point spread market during the 2020-21 season, which was significantly impacted by the coronavirus pandemic. The virus caused chaos throughout the sport, which featured empty stadiums, infectious outbreaks, and scheduling disruptions. These uncertainties created an information gap, particularly regarding player availability, for sports bettors who utilize these types of data to build their models. As such, this market coupled with the pandemic backdrop provides an optimal setting to test the efficiency of herd behavior. Results indicated that the market was statistically inefficient, and from a herding standpoint, contrarian strategies resulted in win percentages that exceeded 50% for various probability cutoffs. This research presents applicable findings related to psychological factors that influence investor behavior, which can be generalized to markets within and outside of sport.

Citation: Alexander Traugutt and Jarid Morton, (2022) "Is herding efficient? Evidence from the college football point spread market", *Economics Bulletin*, Volume 42, Issue 3, pages 1673-1680
Contact: Alexander Traugutt - atraugutt@elon.edu, Jarid Morton - jarid.morton084@topper.wku.edu.
Submitted: January 24, 2022. **Published:** September 30, 2022.

1. Introduction

Herd behavior is a well-known phenomenon in which individuals make decisions based on the actions of others. Instead of relying on rational forecasts or private information, individuals overvalue the actions of others under the assumption that their information is superior. From a financial perspective, a herding-based investment approach can result in significant monetary ramifications, either positive or negative, given the subjective nature of information.

The dynamics of herd behavior have been extensively tested in various financial markets, and a clear path exists to apply the theory in a sport wagering context. This paper focused on the college football point spread market during the 2020-21 college football season, which was significantly impacted by the coronavirus pandemic (Covid-19). The virus caused chaos within the ranks of the sport, which featured empty stadiums, virus outbreaks, and scheduling disruptions. Notably, player availability and roster construction were seemingly in a constant state of flux. As such, average bettors may rely on sharps, or professional bettors, when making their wager selections, despite their own information, which may have led them to place a different wager. The college football point spread market, coupled with the pandemic backdrop, provides an optimal setting to test the potential influence of herd behavior.

2. Literature Review

2.1 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) postulates that all publicly available is included in the price of a security so that consistent profits cannot be achieved (Fama 1970). Thus, investors cannot leverage publicly available information to generate consistent positive returns (Wever and Aadland 2012). Unlike financial markets, where an investment's value is never definitively realized, sports wagering markets provide a clear start and endpoint. As such, these markets provide an optimal setting for testing financial theories and market efficiency (e.g., Gray and Gray 1997; Shank 2017; Sung and Tainsky 2014).

Studies focused on college football markets have not been as widely conducted. Sinkey and Logan (2013) found statistical inefficiencies in the pricing of simple betting strategies by eliminating common behavioral strategies. More specifically, they found statistical inefficiencies in the pricing of home teams, favorites, and those playing weaker competition (Sinkey and Logan 2014). The authors further concluded that the response of sportsbook to bettors' biases are a source of market inefficiency. In particular, as these bookmakers attempt to eliminate profitable opportunities based on behavioral strategies, they, in turn, create profitable avenues for bettors to exploit using simpler betting systems. Bennett (2020) further tested the efficiency of the college point spread market by placing wagers on teams that performed well and performed poorly against the spread in the previous game. Overall, results indicated deviations from market efficiency, particularly for games featuring two non-BCS/Power 5 teams. The market was efficient for games featuring two BCS/Power 5 teams.

2.2 Herd Behavior

Since much of the betting market, as a whole, is made up of recreational bettors (i.e., squares), it is realistic to assume that they would be influenced, to some extent, by professional bettors (i.e., sharps) who have long track records of success. As such, a potential explanation for findings of market inefficiency in wagering markets is the behavioral economic theory of herd behavior, which can explain irrational behavior in various constructs. From finance to marketing, this theory can be utilized to better explain and understand human decision-making. In particular, the behavior theorizes that when an individual observes the choice of people before them, and they rationalize that the people before them made the optimal choice, they will do what other people are doing regardless of their own information (e.g., Banerjee 1992; Bikhchandani *et al*. 1992; Bernheim 1994). As a practical example, when deciding what restaurant to eat at, a person is given two options, and they both have an unknowable optimal outcome of quality. Herd behavior explains that this person will be influenced to choose the restaurant with the longer line out the door instead of the restaurant with no line because the overwhelming number of people waiting at the one restaurant before them implies the level of quality over the other restaurant with no line. They make this decision regardless of their own information about the two restaurants (e.g., Banerjee 1992).

In financial markets, herd behavior can be described as basing investment decisions on the actions of others instead of your own rational information and forecasts. One study has specifically investigated the influence of herd behavior on financial outcomes in the National Football League's (NFL) point spread market. Wever and Aadland (2012) found an overabundance of bets placed on elite NFL teams due, in part, to the increased publicization and media attention. This led to the underpricing of large underdogs resulting in significant profitable returns. The authors posit that their results may be driven by the public's desire to herd toward favored teams because of the intense media coverage they receive in the days leading up to the contest. Further research related specifically to market efficiency and herd behavior in sports wagering markets has yet to be conducted. Thus, this paper will attempt to better understand the influence of herd behavior in a market that has received significantly less research attention.

3. Data

The data for this study were obtained from two sources and included all games, regular season and bowl, from the 2020-2021 season. All game-specific information was compiled from a historical database on sportsbookreviewsonline.com. The public betting information was obtained exclusively from the PRO report on actionnetwork.com, a website dedicated to providing sports betting insights and analytics via manual entry. The Action Network site removes all public betting information following the completion of a predetermined amount of weeks/point in the season. As such, historical information predating the 2020/21 season was unavailable on their primary site.

Consistent with previous studies (e.g., Sung 2020), all push bets were removed from the sample. Such outcomes occur when the score differential between two teams is equal to the point spread. In total, ten games were removed due to this distinction. Additionally, all pick 'em games, which occur when the point spread is set at zero, were excluded from the analysis (e.g., Gray and Gray 1997). The total number of games removed based on this distinction was eight.

The final data set featured a sample size of 531 games, including regular season and bowl contests. Table I provides a detailed overview of each variable.

Table I. Description of variables

Variable	Description
Closing Spread	The closing point spread from the vantage of the home team.
Spread Favorite	1 if the home team was favored based on the spread, 0 if otherwise.
% Money on Spread_Home	Percentage of money placed on the spread from the vantage of the home team.
% Bets on Spread_Home	Percentage of bets placed on the spread from the vantage of the home team
Home Sharp Action*	Presence of sharp money on the home spread. 1 if there was sharp action, 0 if other.
Visitor Sharp Action*	Presence of sharp money on the visitor spread. 1 if there was sharp action, 0 if other.
Home Team Top 25**	Home team ranking. Those within the Top 25 were classified accordingly, all others were assigned a value of 26.
Visitor Team Top 25**	Visitor team ranking. Those within the Top 25 were classified accordingly, all others were assigned a value of 26.
Home Team Talent Composite Ranking***	Measure of home team's talent based on roster construction. Scale: 0 – 1000.
Visitor Team Talent Composite Ranking***	Measure of visiting team's talent based on roster construction. Scale: 0 – 1000.

* Sharps, as identified by the Action Network, are bettors with long track records of success. Sharp action is determined by the website administrators based on the activity of these individuals.
** Weekly Top 25 rankings were compiled from the Associated Press.
*** Team Talent Composite Rankings are algorithmically generated by 247Sports and calculated using prospect rankings and ratings from public media sources.

4. Methodology

Various model specifications have been utilized to model statistical efficiency in the point spread betting market (e.g., Gandar *et al.* 1988; Gray and Gray 1997; Woodland and Woodland 2015). The primary difference in modeling lies in the treatment of the dependent variable as either continuous or dichotomous. For this study, a binary logistic regression model was estimated to determine the degree to which the probability of the home team covering the spread changes, given a set of predictor variables. The first model, hereafter referred to as the basic model, included only general information readily available to all members of the public.

$$W_i = \beta_0 + \beta_1 Closing\ Spread + \beta_2 SpreadFavorite + \beta_3 Home\ Top\ 25 + \beta_4 Home\ Team\ Talent\ Composite\ Ranking + \beta_5 Visitor\ Top\ 25 + \beta_6 Visitor\ Team\ Talent\ Composite\ Ranking \quad (1)$$

The binary outcome W_i takes the value of 1 if the home team covers the spread and 0 if otherwise. The second model, hereafter referred to as the herd behavior model, included all of the variables from the basic model and % Money on Spread, % Bets on Spread, Home Sharp Action, and Visitor Sharp Action. These are the primary variables of interest, given the theoretical foundation of this study.

$$\begin{aligned} W_i = \beta_0 + \beta_1 ClosingSpread + \beta_2 SpreadFavorite + \beta_3 \%MoneyonSpread\_Home \\ + \beta_4 \%\ Bets\ on\ Spread\_Home\ + \beta_5\ Home\ Sharp\ Action \\ + \beta_6 Home\ Top\ 25 + \beta_7 Home\ Team\ Talent\ Composite\ Ranking \\ + \beta_8 Visitor\ Sharp\ Action + \beta_9 Visitor\ Top\ 25 \\ + \beta_{10} Visitor\ Team\ Talent\ Composite\ Ranking \end{aligned} \tag{2}$$

Each of these variables is influenced by public action; that is, their magnitude is increased (decreased) based on the actions of the public. Similar to betting outcomes, it was appropriate to include only the public percentage betting information for one side, given that the visitor value would be the inverse of the home value. Moreover, including the visitor information would lead to issues with multicollinearity. Visiting team variables were included for specific variables unique to the team (e.g., Top 25 ranking).

The EMH postulates that betting lines are inclusive of all publicly available information. Thus, opportunities for sustained above-average returns are not possible. To confirm that the college football point spread market is efficient, the EMH necessitates that:

$$H_0 = \beta_i \ = \ 0\ (i \ = \ 1, 2, 3, \dots) \tag{3}$$

The restrictions of the null hypotheses imply that all coefficients are statistically equal to 0. Any statistically significant deviation from 0 would imply that all publicly available information is not accounted for within the closing line. Thus, opportunities for economic success may be present, which runs counter to the notion of market efficiency, though not a central focus of this study.

Based on the herd behavior model results, predicted probabilities were calculated for each game, indicating the probability of the home team covering the spread. Various probability cutoffs were then established to test whether the outcomes produced by the model aligned with their implied frequency of success (e.g., Gray and Gray 1997).

5. Results & Discussion

Table II provides summary statistics for the variables utilized in the analysis. Of interest from this table is the distribution of values for the variables, which gauged the percentage of bets on the spread and the percentage of dollars wagered. Many individuals will utilize this information to inform their betting strategies, particularly when utilizing a contrarian philosophy (e.g., following the herd or betting against it). As Table II shows, however, mean values for both home and visiting teams were roughly 50%, indicating that the market is quick to correct itself when a particular side is receiving a significant betting share.

Table II. Descriptive statistics

Variable	Minimum	Maximum	Mean	Standard Deviation
Spread Favorite_H	0	1	.41	.49
Spread Favorite_V	0	1	.60	.49
Spread Close_H	-50.5	34.5	-4.54	15.15
Spread Close_V	-34.5	50.5	4.54	15.15
%Bets Spread_H	20	94	49.96	11.04
%Bets Spread_V	6	80	49.99	11.06
%Money Spread_H	6	99	48.61	19.74
%Money Spread_V	1	94	51.43	19.71
Sharp Spread_H	0	1	.41	.49
Sharp Spread_V	0	1	.38	.49
Top 25_H	1	26	22.71	6.77
Top 25_V	1	26	23.19	6.42
TCR_H	71.92	990.52	560.92	201.32
TCR_V	12.53	990.47	533.42	229.90

N = 531. Note. H and V denote Home and Visitor, respectively. TCR is an acronym for Team Composite Ranking.

The results from the basic logistic regression model are provided in Table III. None of the variables are statistically significant or approaching significance. Moreover, the model correctly predicted only 51.2% of the outcomes, which does not meet the generally accepted 52.38% win percentage needed to achieve positive returns.

Table III. Basic model – Logistic regression parameter estimates

Variable	β	p-value	Exp(β)
Spread Close_H	.011	.317	1.011
Spread Favorite_H	.136	.635	1.146
Top 25_H	-.009	.573	.991
TCR_H	.000	.747	1.000
Top 25_V	.010	.557	1.010
TCR_V	.000	.936	1.000
Constant	-.152	.833	.859
Correct Predictions: 51%			

*N = 531. * indicates significance at the .05 level.*
Note. H and V denote Home and Visitor, respectively. TCR is an acronym for Team Composite Ranking.

Results from the herd behavior model Equation (2) are reported in Table IV. As evidenced by the output, only one variable, the percentage of bets on the spread relative to the home team, was a significant predictor at the .05 level. This would indicate that wagering on home teams who are carrying a substantial percentage of the total bet volume would reduce the probability of winning a spread bet placed on the home team. While the remaining variables were not significant predictors of the .05 level, the logistic model correctly predicted 57.1% of betting outcomes during the sample year. It should also be noted that eight of the 11 variables, including the constant, deviated from the market efficiency metric of 0. As such, the market can be considered statistically inefficient given the stringent test implied by Equation (3).

Table IV. Herd behavior model – Logistic regression parameter estimates

Variable	β	p-value	Exp(β)
Spread Close_H	.002	.842	1.002
Spread Favorite_H	.314	.294	1.370
%Money Spread_H	.000	.940	1.000
%Bets Spread_H	-.030	.019*	.971
Sharp Action_H	-.329	.177	.719
Top 25_H	-.006	.738	.994
TCR_H	.000	.910	1.000
Sharp Action_V	-.236	.330	.790
Top 25_V	.014	.388	1.015
TCR_V	.000	.734	1.000
Constant	1.173	.179	3.232
Correct Predictions: 57%			

*N = 531. * indicates significance at the .05 level.*
Note. H and V denote Home and Visitor, respectively. TCR is an acronym for Team Composite Ranking.

To further examine the logit model, comparisons were made between specified probability cutoffs and actual success rates. For example, we first took all games in which the logit model generated predicted probabilities under 45% and then calculated the proportion of successes from those games. Table V displays the complete results. It is evident that the frequencies are closely aligned. In particular, the outcomes associated with probability cutoffs at or above 50%, which accounted for 271 total games or roughly half of the sample, featured win percentages that either approached or exceeded 50%. These results further highlight the accuracy of the model.

Table V. Herd behavior model probabilities versus actual success rate

Logit Probability	Total Games	Games Won (Win %)
$p < 45\%$	117	47 (40%)
$45\% \le p < 50\%$	143	63 (44%)
$50\% \le p < 52.5\%$	91	52 (57%)
$52.5\% \le p < 55\%$	74	36 (49%)
$55\% \le p < 57.5\%$	49	29 (59%)
$p \ge 57.5\%$	57	37 (65%)
Correct Predictions: 50%		

The next analysis is meant to examine the influence of herd behavior, based specifically on the herd behavior logit model results. As indicated by the herd behavior model (Table IV), the percentage of bets on the home team was a statistically significant negative predictor. As such, a contrarian approach of wagering on the visiting team when the percentage of bets for a particular game was high in relation to the home team may uncover an interesting inefficiency in the marketplace. Table VI displays the complete results from this strategy, which support the conclusions drawn from the logit model output. Wagering on the visiting team to cover the spread, particularly in those games where the percentage of bets on the home team was over 60% (which was 20% of the sample), would have resulted in a win percentage of 60%.

Table VI. Betting strategy results

% Bets on Home Team	Total Games	Games Won (Win %)
$p < 45\%$	170	73 (43%)
$45\% \leq p < 50\%$	91	50 (55%)
$50\% \leq p < 55\%$	98	49 (51%)
$55\% \leq p < 60\%$	73	34 (47%)
$p \geq 60\%$	99	59 (60%)
Correct Predictions: 50%		

6. Conclusion and Future Research

The influence of herd behavior on the college football betting market during the pandemic season is evident. Similar to other wagering markets, the results indicated that the market deviated significantly from statistical efficiency. In addition, there were clear opportunities for bettors to utilize a simple strategy of going against the herd, as measured by the percentage of bets on the home team, to achieve win percentages that exceeded 50% for various probability cutoffs.

These findings can serve as the foundation for studies seeking to explore this market and phenomenon further. While this study focused solely on the college football season impacted by the pandemic, utilizing a similar approach for seasons in which the pandemic was not a significant factor would allow for comparisons to be made, particularly as it relates to the empirical impact of herd behavior. In addition, the current approach assessed only statistical efficiency. A larger dataset with odds-related information would also allow for the partitioning of data into training and test sets, which could then be used to analyze economic efficiency. The opportunity to apply other behavioral theories to this market is also clear and would contribute to the emerging field of behavioral economic inquiry relative to sports wagering markets.

References

Banerjee, A. V. (1992) "A Simple Model of Herd Behavior" *The Quarterly Journal of Economics* 107*,* 797-817.

Bennett, R. W. (2020) "Next Game Reaction to Mispriced Betting Lines in College Football" *Applied Economics Letters* 28, 1-4.

Bernheim, B. D. (1994) "A Theory of Conformity" *Journal of Political Economy* 102, 841–877.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992) "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades" *Journal of Political Economy* 100, 992-1026.

Fama, E. F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work" *The Journal of Finance* 25, 383-417.

Gandar, J., Zuber, R., O'Brien, T., and Russo, B. (1988) Testing Rationality in the Point Spread Betting Market" *The Journal of Finance* 45, 995-1008.

Gray, P. K., and Gray, S. F. (1997) "Testing Market Efficiency: Evidence from the NFL Sports Betting Market" *The Journal of Finance* 52, 1725-1737.

Shank, C. A. (2017) "Is the NFL Betting Market Still Efficient?" *Journal of Economics and Finance* 42, 818-827.

Sinkey, M., and Logan, T. (2013) "Does the Hot Hand Drive the Market? Evidence from College Football Betting Market" *Eastern Economic Journal* 40, 583-603.

Sung, Y. T. (2020) "The National Football League and Efficient Market Hypothesis: An Analysis of Trends in Thursday Night football and Post-Bye Week Games" *International Journal of Sport Finance* 15, 69-82.

Sung, Y. T., and Tainsky, S. (2014) "The National Football League Wagering Market: Simple Strategies and Bye Week-Related Inefficiencies" *Journal of Sports Economics* 15, 365-384.

Wever, S., and Aadland, D. (2012) "Herd Behaviour and Underdogs in the NFL" *Applied Economics Letters* 19, 93-97.

Woodland, L. M., and Woodland, B. M. (2015) "The National Football League Season Wins Total Betting Market: The Impact of Heuristics on Behavior" *Southern Economic Journal* 82, 38-54.