

# Beating the House: Identifying Inefficiencies in Sports Betting Markets

Sathya Ramesh\*  
Federal Reserve Board  
sathya.p.ramesh@frb.gov

Ragib Mostofa\*  
Rice University  
rm48@rice.edu

Marco Bornstein\*  
University of Maryland  
marcob@umd.edu

John Dobelman\*  
Rice University  
dobelman@rice.edu

October 22, 2019

## Abstract

Inefficient markets allow investors to consistently outperform the market. To demonstrate that inefficiencies exist in sports betting markets, we created a betting algorithm that generates above market returns for the NFL, NBA, NCAAF, NCAAB, and WNBA betting markets. To formulate our betting strategy, we collected and examined a novel dataset of bets, and created a non-parametric win probability model to find positive expected value situations. As the United States Supreme Court has recently repealed the federal ban on sports betting, research on sports betting markets is increasingly relevant for the growing sports betting industry.

**Keywords:** Sports Betting Markets, Casinos, Win Probability Forecasting

---

\*We would like to thank David Scott, Kathy Ensor, Michael Gordy, Benjamin Gardner, and seminar participants from Rice University and the University of Connecticut for their suggestions and encouragement. Additionally, we thank the Center for Computational Finance and Economic Systems at Rice University for research funding and support. The opinions expressed here are our own, and do not reflect the views of the Board of Governors or its staff. Send correspondence to Sathya Ramesh, Risk Analysis Section, Board of Governors of the Federal Reserve System, 20th & C Streets, NW, Washington, DC 20551, Telephone: (202) 973-6124, e-mail: sathya.p.ramesh@frb.gov

# 1 Introduction

Casino games are designed to make gamblers choose from a range of negative expected value outcomes. In the short run, a gambler may profit; however, in the long run, the house always wins due to the Law of Large Numbers. The sentiment of the house inevitably winning is paralleled in financial markets by the strong form of the Efficient Markets Hypothesis (Fama 1970), which states that it is impossible to outperform the market due to the informational efficiency of markets.

While there are some similarities in market structure between financial markets and casino sports books, there are also several key differences (Levitt 2004). In financial markets, price is dictated by supply and demand. However, in sports betting markets, each casino dictates the “price” (odds ratio) at which the “stock” (specific bet) will trade. Levitt stresses the importance of the casino providing the right price for each bet by mentioning that the bookmaker can be exposed to significant losses if the price set by the bookmaker is incorrect. Another key difference between financial markets and sports betting markets is that the Securities and Exchange Commission requires brokers to execute trades at the National Best Bid and Offer, whereas it is up to the individual bettor to discover which casino is offering the best risk-reward ratio.

As Moskowitz (2015) notes, several papers<sup>1</sup> study the efficiency of sports betting markets, with the evidence somewhat mixed. We define an inefficient sports betting market as a market in which a betting strategy consistently generates profit. (Snyder 1978; Sauer et al 1988; Constantinou et al 2013). From our analysis, we find that there are exploitable inefficiencies due to differences in price offerings by individual casinos. Unlike any of the aforementioned papers (which tend to focus on one or two sports), we utilize a single pricing model to identify mispriced bets across several sports (namely the NFL, NBA, NCAAF, NCAAB, and WNBA). We extend Levitt’s conclusion that bookmakers are more skilled at predicting game outcomes than bettors to the other major sports that we examine in this paper. This assumption that bookmakers are the most skilled at predicting game outcomes is key to formulating our model as we chiefly use information and data gathered from the entire market of casinos to take advantage of the mispricing by inefficient casinos. Due to our successful identifications of mispricings in sports betting markets, we were able to build a betting strategy that consistently generates profit across several different sports.

---

<sup>1</sup> Snyder (1978), Vergin and Scriabin (1978), Ali (1979), Losey and Talbott (1980), Hausch et al (1981), Asch et al (1984), Zuber et al (1985), Thaler and Ziemba (1988), Sauer et al (1988), Gandar et al (1988), Camerer (1989), Golec and Tamarkin (1991), Woodland and Woodland (1994), Dare and MacDonald (1996), Gray and Gray (1997), Avery and Chevalier (1999), Kuypers (2000), and Lee and Smith (2002) have commented on the efficiency of various sports betting markets.

## 2 Terminology

Before diving deeper into the subject matter, we find it essential to define some of the parlance involved with sports betting.

### 2.1 Point Spread & Spread Bets

To paraphrase Stern (1991), the perceived difference between two sporting teams is measured by the point spread. For example, if Team A is determined to have a point spread of -6.5 when matched up against Team B, then Team A is favored to beat Team B by 6.5 points. In this scenario, Team A is the favorite (as they are favored to win) and Team B is the underdog. A bet placed on Team A will be a winning bet if and only if Team A beats Team B by more than 6.5 points. Conversely, Team B will have an inverse point spread of +6.5 points. A bet placed on Team B will be a winning bet if and only if Team B either wins *or* loses by less than 6.5 points. The betting in this scenario is an example of a spread bet. A spread bet is designed to have even odds, meaning that chance of success and failure are equal at 50%. For this reason, a spread bet generally pays out 91 cents for every dollar wagered if the bet is successful. The casino collects the other 9 cents, which is called a “vigorous”. If there is an equal amount of money on either side, then the casino collects 9 cents for each dollar risked, and is exposed to no risk.

Spreads can also be understood in the following manner, where  $PFS$  represents the Projected Final Score of each team.

$$\text{Point Spread} = PFS_{\text{Favorite}} - PFS_{\text{Underdog}} \quad (1)$$

Spreads often include the half point “hook” in order to avoid ties, as most major sports (basketball, football, soccer, hockey, etc.) do not have fractional point scores. In the event that the spread is a whole number (i.e. Team A is +6), and the outcome is Team A wins by 6 points, money is returned to anyone who bets on the game. This kind of event is called a “push”.

### 2.2 Moneyline Bets

The Moneyline bet is a wager on which team will win. Moskowitz mentions that Moneyline bets pay out various amounts of dollars depending on which team is bet. (Moskowitz 2015). For example, if the moneyline for Team A is -500 and the moneyline for Team B is +300, a successful wager of 500 dollars on Team A nets a 100 dollar profit while a successful wager of 100 dollars on Team B nets a 300 dollar profit. Our algorithm solely focuses on identifying mispriced moneyline bets.

## 2.3 Market Return

Defining average market return in the sports betting context is important since this activity does not lend itself to traditional financial return concepts, such as a comparison to given benchmarks. A bettor’s average return in this market will be equal to the mean return of their wagers, weighted for bet sizes. Table 1 reveals that a bettor who randomly wagers on spread bets will have a mean return of approximately -4.4% for all sports, and a bettor who randomly wagers on moneyline bets will have a mean return between -6.4% and 0.8% for all sports. As 0% is included in all of the randomized moneyline bet intervals for all sports, we view the average market return as equal to 0%. With this definition of average market return, we define above average market return for sports betting to be long run positive returns. Table 1 will be discussed in further detail in the Results section.

## 3 Model & Data

### 3.1 Probability and Expected Value

There are 16 casino sportsbooks in Las Vegas that we obtained data from. For any given sports game, at most 16 casinos will provide point spreads and moneyline betting odds. The point spread and moneyline odds vary from casino to casino. Thus, from our obtained data, we end up with a range of point spreads and moneyline odds. As there are a plethora of teams in College Basketball and College Football, we only considered games where one team was a member of a “Power Five” conference<sup>2</sup>. For all other sports, we considered all games between all teams, excluding pre-season (exhibition) games but including post-season (i.e. playoff) games.

Our algorithm utilizes this range of point spreads to determine the probability of a team winning the game. Unlike Stern, we do not assume that a team’s implied probability of winning a game from the point spread is normally distributed. One limitation of Stern’s model is its ability to only price games with the absolute value of the spread less than six . Our model is non-parametric and allows us to price moneyline bets for any point spread. The following equation describes our Bayesian formula for a team’s likelihood to win a game, at point spread PS.

$$P(\text{Win} \mid \text{PS}) = \text{Historical Win \% of Teams with Point Spread} = \text{PS} \quad (2)$$

This formula is the engine of our algorithm. We theorize that a team at point spread PS will win at

---

<sup>2</sup> The Power Five Conferences are defined to be the Atlantic Coast Conference, Big Ten Conference, Big 12 Conference, Pac-12 Conference, and Southeastern Conference.

the same rate at which historical teams at spread PS have won. If we know the probability of an event, and we know the payout for making a correct wager, we can define expected value for a one dollar bet as follows.

$$\text{Expected Value} = [\text{P(Winning)} * \text{Payout}] - \text{P(Losing)} \quad (3)$$

The expected value of a one dollar bet is defined as the probability of winning multiplied by the payout for winning, subtracted by the probability of losing multiplied by the money wagered. Since the money wagered in a one dollar bet is 1, the expected winnings are now solely subtracted by the probability of losing. This is how Equation 3 is formed.

The expected value of a bet symbolizes the expected money made or lost from the bet. Theoretically, selecting bets that have a positive expected value, and betting on them, would generate profit over time. This is the foundation of our betting algorithm.

Clearly, calculating the probability and expected value requires the spreads and payouts from Vegas casinos and historical spread data and results (i.e. which teams won and lost). Vegasinsider.com has spread and payout data from Vegas casinos for many sporting events, as well as the results of these events. Killersports.com has historical spread data and results.

### 3.2 Algorithms

As mentioned earlier, a range of point spreads and moneylines are obtained from vegasinsider.com. As we assume that the bookmakers know how to predict game outcomes best and are incentivized to provide accurate prices (resulting in the market as a whole being informationally efficient), we prefer to obtain the most updated spreads and moneylines. Our data analysis agreed with this assumption, and revealed that bookmakers' last update (i.e. whether a player was injured, etc.) generally occurred within an hour of the game. From this finding, we decided that the casino's last update was the most accurate indicator of the various casino specified variables (point spread and moneyline payout), as it was closest to game time and thus contained the most recent information.

We have two slightly different strategies that utilize the range of spreads to find the probability of winning, which we will refer to as "Simple" (Algorithm 1) and "Weighted" (Algorithm 2). Both algorithms utilize an input mapping of spreads to their respective probabilities, but the Weighted algorithm also uses an input mapping of spreads to their respective frequencies. The simple algorithm returns the probability of a team winning via calculating the simple average of the probabilities of the unique spreads. The weighted algorithm returns the probability of a team winning via calculating the weighted average of the probabilities of the spreads by using the frequency of the spread as the weighting factor. Below, Algorithm 1 and Algorithm

2 show the pseudocode of each algorithm.

**Place Algorithm 1 Here**

**Place Algorithm 2 Here**

With the probability determined using either strategy, and the payout determined as the maximum of the moneyline payouts, we can utilize the expected value formula in Equation 3 to find the expected value of picking a team to win a game. If our probability calculations are correct, and thus our expected value computations, we should make money by wagering on bets with positive expected values.

### 3.3 Epsilon and Expected Value Thresholds

As the returns in Table 2 will demonstrate, simply betting on all games with positive expected value generates positive return on investment for three out of five sports when the simple algorithm is used to calculate probability of winning, and five out of five sports when the weighted algorithm is used to calculate the probability of winning. This finding implies that there are inefficiencies in the market, as there exists a method of obtaining positive returns. For reference, randomized betting returns on the same data are provided in Table 1. Both of these tables will be analyzed in further detail in the Results section of the paper. From data analysis, we were able to make improvements on our model.

The first improvement on our model regards the issue of predicting games with large spreads. This scenario is why Stern’s model works best pricing bets with the absolute value of the spread less than six, as the tails are hard to predict. If the spread is large, it implies that one team is heavily favored to beat the other team, and our model lacks the historical sample size at such large spreads to create a realistic probability estimate. Additionally, casinos make the payouts for betting on underdogs in this scenario ludicrously high, resulting in the model generating a high expected value for underdogs in these cases.

To account for the effect of these games in the tails, we created the “Epsilon Threshold”. The goal of the Epsilon Threshold is to provide a symmetric set of bounds for the probabilities of teams winning games that we are willing to bet on when the bet has positive expected value. For bets outside the bounds, we always go with the probability favorite. After determining which Epsilon Value yields the highest total return on investment, we can define the optimal bettable range of winning percentages as:  $(.5 - \epsilon, .5 + \epsilon)$ . The optimal Epsilon value varies significantly from sport to sport.

The second improvement we designed is the “Expected Value Threshold”. Expected Value Threshold is the the value of Expected Value that a bet must be greater than in order to place the bet. The default

Expected Value Threshold that we use is 0. In this default case, any bet that has a positive Expected Value will be bet on. In general, as the Expected Value Threshold increases so does the return on investment, which leads us to believe that the model is detecting large inefficiencies very well.

### 3.4 Betting Algorithm

The Betting Algorithm (Algorithm 3) decides which team to bet on by examining the expected value of the bet and applying the epsilon and expected value thresholds. The algorithm takes the following inputs: the expected value of betting on the favorite, the expected value of betting on the underdog, the probability that the favorite wins, the probability that the underdog wins, the value of the epsilon hyper-parameter, and the value of the expected value threshold. The output of the algorithm is a decision to bet on the favorite, underdog, or neither team depending on the filtering criteria of expected value, epsilon threshold, and expected value threshold. Below, Algorithm 3 shows the pseudocode of the algorithm.

#### Place Algorithm 3 Here

Return on Investment (ROI %) is computed outside of the algorithm by comparing the algorithm's bets to the real world outcomes. If the algorithm picks a team that wins, the winnings from that bet are equivalent to the payout of the correct bet. Else, if the algorithm picks a team that loses, the losses from that bet are negative one (as every bet is a one dollar bet). Else, if the algorithm doesn't pick a team, there is no bet, so there are no winnings or losses. The ROI % is the one hundred times sum of the winnings and losses divided by the number of bets.

## 4 Results

In order to provide benchmarks for our algorithm returns, we will first discuss Table 1, which displays the returns from randomly placing spread and moneyline bets. Next, we will contrast these results with the findings from Table 2, which displays the returns for our two win probability models: Simple and Weighted. Finally, we will showcase a year by year, sport by sport breakdown of the optimal ROI with 2

#### Place Table 1 Here

Table 1 summarizes returns from three different strategies of randomly betting on games. Algorithms 4 and 5 describe the pseudocode that implements each of these strategies. The confidence intervals in each of the panels are 95th percentile confidence intervals, and each strategy involves betting on every game in

the sample data. From Panel A, it's evident that randomly picking spread bets (regardless of sport) returns a mean ROI of around -4.4%. In this randomization approach, there is an equal chance the strategy will choose either team to cover the spread. In Panel B, the mean ROIs have more variation than Panel A, and the mean returns are close to 0% (excluding NCAAB). In the randomization approach used to generate Panel B, the favorite and the underdog have an equal chance to be picked to win the game. In Panel C, the mean ROIs are similar to Panel C's, but generally smaller (excluding NCAAB). The randomization approach used to generate Panel C assigns a 67% chance that the favorite will be chosen to win the game, and a 33% chance that the underdog will be chosen to win the game.

From comparing all three panels, it's clear that randomly placing moneyline bets generally have a larger returns on investment than randomly placing spread bets. From these results, we chose to build our betting strategies around successfully pricing and placing moneyline bets. Additionally, it's notable that all of the 95% confidence intervals (excluding NBA spread bets from Panel A) contain 0 in the interval, as we have defined an inefficient market as a market where a strategy can consistently generate profit. Clearly, we cannot reject the null hypothesis that sports betting markets are jointly efficient for all of these sports for any kind of bet.

### Place Table 2 Here

This table summarizes returns from two win probability models: Simple and Weighted. Algorithms 1 and 2 describe the pseudocode that implements each of these models. Algorithm 3 describes how these probability models are incorporated into the betting process. Both Panel A and B follow the same structure. First, returns are reported for bets made sheerly on an expected value basis derived from the specific win probability model. Next, returns are reported for the betting algorithm with the Epsilon Threshold parameter. Finally, returns are displayed for the betting algorithm augmented with the Expected Value Threshold parameter as well as the Epsilon Threshold.

Returns always increase at each step, as we have selected the optimal values of the Epsilon and Expected Value Thresholds. We found these optimal parameter values by maximizing Total Return, which is defined as  $ROI * N$ , where  $N$  is the number of bets. In Panel A, the returns resulting from strictly making positive expected value bets according to the Simple model vary across sports, with the NCAAF and NCAAB having negative ROI. In comparison, the returns from making positive expected value bets according to the Weighted model also vary across sports, but are always positive. As aforementioned, returns increase when more parameters are included. The best conclusion to draw from comparing the returns augmented with optimal Epsilon and EV Thresholds for both algorithms is that neither model really dominates the other.



This leads us to believe that choosing Simple or Weighted on a sport by sport basis is probably the best idea. Next, we'll do a year by year breakdown of each sport's ROI to provide deeper detail and context for the results.

### **Place Table 3 Here**

Table 3 shows a year by year, sport by sport breakdown of optimal ROI (bets made by betting algorithm augmented with optimal Epsilon and Expected Value Threshold parameters), with the S&P 500 ROI listed as a benchmark. The All Leagues column shows a weighted ROI generated by summing the Total Return ( $\text{ROI} * \text{Number of Games}$ ) for each sport, and then dividing by the total number of games bet upon for each year. We do not benefit from compounding, as we designate each bet as an independent \$ 1 bet. Additionally, even though we are computing the necessary inputs, we do not utilize Kelly's criterion.

The sample of games we have excludes the early 2000's, 2008, and 2018, which are the years in the 21st century in which S&P 500 had a negative return. In general, the S&P 500 outperforms the sports betting returns (excluding 2012 for Simple, and 2011 and 2015 for both strategies). We present this table as the foundation of the argument that sports betting should be viewed as an alternative asset management strategy, as it logically has no relation to the stock market (market neutral). The next table will show the yearly sample size, and we will explain why we are unable to collect data for more past years.

### **Place Table 4 Here**

This table supports Table 3, and provides the sample size of bets for each sport in each year. In general, as time goes on, we bet on more games. This is a feature of the data as we had to web scrape most of the line movements of the games from an internet archive ([archive.org/web](http://archive.org/web)) as the site the data were originally found at ([vegasinsider.com](http://vegasinsider.com)) does not display line movement data for any games outside of the current year, and it is challenging to find older line movement data. We are also limited by our database of line movements. In certain sports (WNBA), sufficient historical spread win probability data is lacking due to the sparsity of both games played and line movements measured. An interesting result of this comparison of tables is that the sample size of the Simple win probability model tends to be considerably larger for more recent years of data.

Comparing the results from Table 1 and Tables 2 and 3 highlights that the optimal ROI's from Table 2 are much larger than the mean ROI from Table 1. To show that the expected optimal ROI's are not a feature of over-fitting the hyper-parameters (Epsilon Value and Expected Value Threshold) to the sample data, we utilized the bootstrap procedure. If we can show that the expected optimal ROI's for the betting strategy

are larger than zero with statistical significance, we have proved that the strategy consistently produces a profit. Proving that the strategy consistently produces a profit clearly indicates that the sports betting markets are inefficient.

## 4.1 Bootstrap Results

In order to demonstrate the robustness of our results, we utilized the bootstrap procedure. The algorithm that describes our implementation is Algorithm 6. This algorithm is utilized for each sport, and the number of times the data is resampled with replacement for each sport is equivalent to the number of sample games for that sport. 10,000 iterations of this algorithm are used to generate the desired bootstrap samples from which Optimal ROI, Optimal Epsilon Value, Optimal Expected Value Threshold, and Optimal Number of Games can be computed. Univariate and bivariate visualizations of the aforementioned variables in the bootstrap samples are available in the Online Supplement. The univariate visualizations are histograms with optimal bin sizing calculated by Scott’s Rule (Scott 1979), and the bivariate visualizations are contour plots with densities calculated via averaged shifted histograms (Scott 1985). Both visualization techniques are non-parametric.

From the bootstrap samples, Table 5 contains 95% Confidence Intervals for each variable sampled. From the univariate visualizations in the Online Supplement, it’s evident that some of the samples are clearly skewed. In order to handle the skewness of the data, the confidence intervals are calculated by two different non-parametric procedures: Percentile Intervals (Efron 1993) and High Density Intervals (Hyndman 1996).

### Place Table 5 Here

Table 5 displays two-sided 95% confidence intervals for the bootstrapped data. Panel A and B highlight the intervals for bets relying upon the Simple win probability model, and Panel C and D highlight the intervals for bets relying upon the Weighted win probability model. Both Panel A and C utilize 95% percentile intervals, and Panel B and D utilize 95% high density intervals. All ROI and Epsilon Value intervals for all sports in each panel do not contain zero.

To examine whether the optimal ROI values (or optimal Epsilon Values) for all sports are statistically significant and greater than zero, we must test the following set of joint hypotheses:  $H_0 = 0$  and  $H_A > 0$ . In order to perform the joint hypothesis test (as we are checking these hypotheses for all sports concurrently), we use the Bonferroni correction with  $\alpha = 0.05$ . Since we are testing five sports concurrently with the aforementioned hypotheses for each win probability model, we must now examine whether all of the one

sided 99% confidence intervals contain zero for each win probability model.

### Place Table 6 Here

Table 6 displays one-sided 99% confidence intervals for bootstrapped data. Panel A highlights the intervals for bets relying upon the Simple win probability model, and Panel B highlights the intervals for bets relying upon the Weighted win probability model. All ROI intervals for all sports in each panel do not contain zero. As we are utilizing an one-sided interval, the null and alternative hypotheses for ROI, EV Threshold, and Epsilon Value are respectively:  $H_0 = 0$  and  $H_A > 0$ . By examining the intervals in the table and using the Bonferroni correction (with  $\alpha = 0.05$ ) for each win probability model, we reject the null hypothesis that the optimal returns from this strategy are equivalent to zero for every sport. By again examining the intervals in the table and using the Bonferroni correction (with  $\alpha = 0.05$ ) for each win probability model, we reject the null hypothesis that the optimal Epsilon Value is equivalent to zero for every sport. From the above, we have shown that the optimal ROI is statistically significant and greater than zero, meaning that there is a betting strategy that consistently generates profit.

## 5 Conclusion

Levitt (2004) succinctly states that there is little evidence of bettors who are able to beat the bookies systematically. For horseracing bets, Asch et al (1982) were not able to devise a successful strategy based solely on observable betting odds. Through our analysis, we have shown that we have created a betting algorithm that solely relies upon betting odds and is capable of generating profit for multiple sports concurrently. If the market is efficient, there should not be any skillful way to generate above average market returns in the long run. Through our research and successful algorithm, we demonstrated the existence of market inefficiencies in sports betting markets. Our paper is the first to propose a single theory which identifies inefficiencies across sports through generating positive returns.

We encourage future researchers to further explore the benefits and costs of using compounding bets rather than independent \$1 bets that we used in this paper, as well as properly utilizing Kelly's portfolio growth criterion (Kelly 1956). In addition, we also leave it to future researchers to propose improvements on our win probability models and betting algorithm parameters.

As the Supreme Court of the United States has struck down a 1992 law prohibiting sports betting<sup>3</sup>, we anticipate that betting markets in the United States will continue to grow larger and more popular. Indeed,

---

<sup>3</sup>Murphy v. National Collegiate Athletic Association, 584 U.S. — 2018

several states<sup>4</sup> have already legalized sports betting since the monumental decision. For now, we urge bettors to take advantage of these newfound opportunities and enjoy beating the house.

---

<sup>4</sup>New Jersey, Delaware, Mississippi, and West Virginia are a few examples.

## References

- Ali, M. M., 1979. Some evidence of the efficiency of a speculative market. *Econometrica* 47, 387–392.
- Asch, P., Malkiel, B. G., Quandt, R. E., 1982. Racetrack betting and informed behavior. *Journal of Financial Economics* 10, 187–194.
- Asch, P., Malkiel, B. G., Quandt, R. E., 1984. Market efficiency in racetrack betting. *Journal of Business* 57, 165–175.
- Avery, C., Chevalier, J., 1999. Identifying investor sentiment from price paths: The case of football betting. *The Journal of Business* 72, 493–521.
- Camerer, C. F., 1989. Does the basketball market believe in the “Hot Hand”. *American Economic Review* 79, 1257–1261.
- Constantinou, A. C., Fenton, N. E., 2013. Profiting from arbitrage and odds biases of the european football gambling market. *The Journal of Gambling Business and Economics* 7, 41–70.
- Dare, W. H., MacDonald, S. S., 1996. A generalized model for testing the home and favorite team advantage in point spread markets. *Journal of Financial Economics* 40, 295–318.
- Efron, B., Tibshirani, R., 1994. *An introduction to the bootstrap*. New York: Chapman & Hall.
- 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., Russo, B., 1988. Testing rationality in the point spread betting market. *Journal of Finance* 43, 995–1008.
- Golec, J., Tamarkin, M., 1991. The degree of inefficiency in the football betting market. *Journal of Financial Economics* 30, 311–323.
- Gray, P. K., Gray, S. F., 1997. Testing market efficiency: Evidence from the NFL sports betting market. *The Journal of Finance* 52, 1725–1737.
- Hausch, D. B., Ziemba, W. T., Rubinstein, M., 1981. Efficiency of the market for racetrack betting. *Management Science* 27, 1435–1452.
- Hyndman, R. J., 1996. Computing and graphing high density regions. *The American Statistician* 50, 120–126.
- Kelly, J., 1956. A new interpretation of information rate. *Bell System Technical Journal* 35, 917–926.

- Kuypers, T., 2000. Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics* 32, 1353–1363.
- Lee, M., Smith, G., 2002. Regression to the mean and football wagers. *Journal of Behavioral Decision Making* 15, 329–342.
- Levitt, S. D., 2004. Why are gambling markets organised so differently from financial markets? *The Economic Journal* 114, 223–246.
- Losey, R. L., Talbott, J. C., 1980. Back on the track with the efficient markets hypothesis. *The Journal of Finance* 35, 1039–1043.
- Moskowitz, T. J., 2015. Asset pricing and sports betting. Chicago Booth Research Paper pp. 15–26.
- Sauer, R. D., Brajer, V., Ferris, S. P., Marr, M. W., 1988. Hold your bets: Another look at the efficiency of the gambling market for national football league games. *Journal of Political Economy* 96, 206–213.
- Scott, D. W., 1979. On optimal and data-based histograms. *Biometrika* 66, 605–610.
- Scott, D. W., 1985. Average shifted histograms: Effective nonparametric density estimators in several dimensions. *The Annals of Statistics* 13, 1024–1040.
- Snyder, W. W., 1988. Horse racing: Testing the efficient markets model. *Journal of Finance* 33, 1109–1118.
- Stern, H., 1991. On the probability of winning a football game. *The American Statistician* 45, 179–183.
- Thaler, R. H., Ziemba, W. T., 1988. Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives* 2, 161–174.
- Vergin, R. C., Scriabin, M., 1978. Winning strategies for wagering on National Football League games. *Management Science* 24, 809–818.
- Woodland, L. M., Woodland, B. M., 1994. Market efficiency and the favorite-longshot bias: the baseball betting market. *The Journal of Finance* 49, 269–279.
- Zuber, R. A., Gandar, J. M., Bowers, B. D., 1985. Beating the spread: Testing the efficiency of the gambling market for national football league games. *Journal of Political Economy* 93, 800–806.

Table 1: Randomized Betting Returns (\$1 Bets)

<b>Panel A: Equally Randomized Spread Betting Returns</b>					
	NFL	NBA	NCAAF	NCAAB	WNBA
Total Games Bet	922.00	4019.00	951.00	1140.00	255.00
Mean Return on Investment (ROI) %	-4.35	-4.39	-4.41	-4.39	-4.39
95% ROI Confidence Intervals	(-10.31, 1.49)	(-7.27, -1.43)	(-10.26, 1.39)	(-9.80, 1.08)	(-15.94, 7.27)

<b>Panel B: Equally Randomized Moneyline Betting Returns</b>					
	NFL	NBA	NCAAF	NCAAB	WNBA
Total Games Bet	922.00	4019.00	951.00	1140.00	255.00
Mean Return on Investment (ROI) %	0.75	0.54	0.30	-6.44	-0.18
95% ROI Confidence Intervals	(-6.48, 7.90)	(-3.26, 4.23)	(-8.14, 8.59)	(-12.97, 0.09)	(-14.62, 14.07)

<b>Panel C: Unequally Randomized Moneyline Betting Returns</b>					
	NFL	NBA	NCAAF	NCAAB	WNBA
Total Games Bet	922.00	4019.00	951.00	1140.00	255.00
Mean Return on Investment (ROI) %	0.72	-1.04	-1.52	-3.78	-1.39
95% ROI Confidence Intervals	(-6.22, 7.56)	(-4.47, 2.39)	(-8.49, 5.67)	(-9.48, 1.92)	(-14.35, 12.25)

This table summarizes returns from three different strategies of randomly betting on games. Algorithms 4 and 5 describe the pseudocode that implements each of these strategies. The confidence intervals in each of the panels are 95th percentile confidence intervals, and each strategy involves betting on every game in the sample data. From Panel A, it's evident that randomly picking spread bets (regardless of sport) returns a mean ROI of around -4.4%. In this randomization approach, there is an equal chance the strategy will choose either team to cover the spread. This approach mirrors the fact that spread bets are designed to have an equal chance of the favorite or underdog covering the spread. In Panel B, the mean ROIs have more variation than Panel A, and the mean returns are close to 0% (excluding NCAAB). In the randomization approach used to generate Panel B, the favorite and the underdog have an equal chance to be picked to win the game. In Panel C, the mean ROIs are similar to Panel B's, but generally smaller (excluding NCAAB). The randomization approach used to generate Panel C assigns a 67% chance that the favorite will be chosen to win the game, and a 33% chance that the underdog will be chosen to win the game. From comparing all three panels, it's clear that randomly placing moneyline bets generally have a larger returns on investment than randomly placing spread bets. From these results, we chose to build our betting strategies around successfully placing moneyline bets.

Table 2: Betting Algorithm Returns (\$1 Bets)

Panel A: Simple Algorithm Results					
	NFL	NBA	NCAAF	NCAAB	WNBA
<b>Simple</b>					
Total Games Analyzed	922.00	4019.00	951.00	1140.00	255.00
Games Bet (Simple)	742.00	2386.00	667.00	709.00	226.00
Return on Investment (ROI) %	10.81	3.42	-2.49	-1.10	4.38
<b>Simple with Optimal Epsilon Threshold</b>					
Epsilon Value	0.34	0.38	0.25	0.17	0.20
Bet Games	742.00	2386.00	667.00	709.00	226.00
ROI %	11.55	4.02	2.97	1.39	7.42
<b>Simple with Optimal Epsilon and EV Thresholds</b>					
EV Threshold	0.013	0.010	0.083	0.007	0.084
Games Bet	567.00	1500.00	292.00	645.00	151.00
ROI %	16.57	9.04	7.29	4.07	17.01
Panel B: Weighted Algorithm Results					
	NFL	NBA	NCAAF	NCAAB	WNBA
<b>Weighted</b>					
Total Games Analyzed	922.00	4019.00	951.00	1140.00	255.00
Games Bet (Weighted)	766.00	2400.00	664.00	685.00	235.00
Return on Investment (ROI) %	5.73	0.66	1.60	0.94	4.55
<b>Weighted with Optimal Epsilon Threshold</b>					
Epsilon Value	0.12	0.35	0.25	0.37	0.16
Bet Games	766.00	2400.00	664.00	685.00	235.00
ROI %	8.69	1.57	6.06	2.67	9.70
<b>Weighted with Optimal Epsilon and EV Thresholds</b>					
EV Threshold	0.022	0.040	0.010	0.030	0.023
Games Bet	662.00	619.00	599.00	284.00	216.00
ROI %	9.26	11.18	8.04	13.02	12.98

This table summarizes returns from two win probability models: Simple and Weighted. Algorithms 1 and 2 describe the pseudocode that implements each of these models. Algorithm 3 describes how these probability models are incorporated into the betting process. Both Panel A and B follow the same structure with returns reported for bets made sheerly on an expected value basis derived from the specific win probability model. Next, returns are reported for the betting algorithm with the Epsilon Threshold parameter. Finally, returns are displayed for the betting algorithm augmented with the Expected Value Threshold parameter as well as the Epsilon Threshold. Returns always increase at each step, as we have selected the optimal values of the Epsilon and Expected Value Thresholds. We chose these parameter values to optimize for maximizing Total Return, which is defined as  $ROI * N$ , where  $N$  is the number of bets. In Panel A, the returns resulting from strictly making positive expected value bets according to the Simple model vary across sports, with the NCAAF and NCAAB having negative ROI. In comparison, the returns from making positive expected value bets according to the Weighted model also vary across sports, but are always positive. As aforementioned, returns increase when more parameters are included. The best conclusion to draw from comparing the returns with both algorithms augmented with optimal Epsilon and EV Thresholds is that neither model really dominates the other, which leads us to believe that choosing Simple or Weighted on a sport by sport basis is probably the best idea.



Table 3: Yearly Optimal Return on Investment [%] (\$1 Bets)

**Panel A: Simple Return on Investment**

Year	NFL	NBA	NCAAF	NCAAB	WNBA	All Leagues	Sample Size	S&P 500
2017	34.01	4.20	0.00	-0.30	7.58	3.38	839	21.69
2016	-1.56	8.56	-6.57	-0.22	-3.83	2.22	624	11.80
2015	43.71	17.69	26.12	18.24	107.79	29.09	365	1.34
2014	4.03	2.13	64.00	-13.87	0.00	8.22	164	13.53
2013	-15.16	11.40	-8.82	16.25	100.00	10.64	189	32.21
2012	9.77	25.11	7.22	8.33	77.78	17.51	128	15.84
2011	76.03	16.79	12.00	—	—	29.19	272	2.06
2010	-7.89	6.07	0.85	—	—	2.84	384	14.93
2009	1.10	4.37	11.27	—	—	4.25	190	26.42

**Panel B: Weighted ROI Return on Investment**

Year	NFL	NBA	NCAAF	NCAAB	WNBA	All Leagues	Sample Size	S&P 500
2017	12.46	-3.68	-25.00	3.29	7.38	3.28	465	21.69
2016	5.95	7.57	5.59	-6.74	15.07	5.23	475	11.80
2015	14.67	22.88	10.77	23.95	47.64	18.35	344	1.34
2014	-12.59	7.05	51.40	-10.83	0.00	6.17	128	13.53
2013	-13.31	24.82	-43.14	-2.06	100.00	-0.89	121	32.21
2012	-12.67	36.52	9.34	8.33	5.56	10.00	150	15.84
2011	46.96	23.09	11.05	—	—	26.23	199	2.06
2010	17.60	11.65	6.91	—	—	11.55	318	14.93
2009	-0.20	-2.88	9.19	—	—	2.15	170	26.42

This table shows a year by year, sport by sport breakdown of optimal ROI (bets made by betting algorithm augmented with optimal Epsilon and Expected Value Threshold parameters), with the S&P 500 ROI listed as a benchmark. The All Leagues column shows a weighted ROI generated by summing the Total Return (ROI \* Number of Games) for each sport, and then dividing by the total number of games bet upon for each year. We do not benefit from compounding, as we designate each bet as an independent \$ 1 bet. The sample of games we have excludes the early 2000's and 2008 and 2018, which are the years in the 21st century in which S&P 500 had a negative return. In general, the S&P 500 outperforms the sports betting returns (excluding 2012 for Simple, and 2011 and 2015 for both strategies). We present this table as the foundation of the argument that sports betting should be viewed as an alternative asset management strategy, as it logically has no relation to the stock market (market neutral).

Table 4: Yearly Sample Sizes (Games Bet On)

**Panel A: Simple Algo Sample Size (Games Bet On)**

Year	NFL	NBA	NCAAF	NCAAB	WNBA	All Leagues
2017	16	343	0	354	126	839
2016	145	266	90	114	9	624
2015	107	124	29	93	12	365
2014	56	81	17	10	0	164
2013	14	85	15	73	2	189
2012	33	61	31	1	2	128
2011	60	174	38	0	0	272
2010	72	267	45	0	0	384
2009	64	99	27	0	0	190

**Panel B: Weighted Algo Sample Size (Games Bet On)**

Year	NFL	NBA	NCAAF	NCAAB	WNBA	All Leagues
2017	21	124	2	141	177	465
2016	165	95	158	44	13	475
2015	141	53	84	43	23	344
2014	64	34	27	3	0	128
2013	17	43	25	34	2	121
2012	36	33	79	1	1	150
2011	62	66	71	0	0	199
2010	80	131	107	0	0	318
2009	76	40	54	0	0	170

This table supports Table 3, and provides the sample size of bets for each sport in each year. In general, as time goes on, we bet on more games. This is a feature of the data as we had to web scrape most of the line movements of the games from an internet archive ([archive.org/web](http://archive.org/web)) as the site the data were originally found at ([vegasinsider.com](http://vegasinsider.com)) does not display line movement data for any games outside of the current year, and it is challenging to find older line movement data. We are also limited by our database of line movements. In certain sports, sufficient historical spread win probability data is lacking due to the sparsity of both games played and line movements measured. An interesting result of this comparison of tables is that the sample size of the Simple win probability model tends to be considerably larger for more recent years of data.

Table 5: Two Sided 95% Bootstrap Confidence Intervals

**Panel A: Simple High Density Intervals**

Sport	ROI %	EV Threshold	Epsilon Value	% of Games Bet
NFL	(7.04, 35.81)	(0.00, 0.03)	(0.11, 0.36)	(0.38, 0.82)
NBA	(3.74, 16.00)	(0.00, 0.02)	(0.19, 0.46)	(0.28, 0.60)
NCAAB	(1.36, 16.08)	(0.01, 0.03)	(0.14, 0.37)	(0.29, 0.60)
NCAAF	(1.10, 16.00)	(0.00, 0.08)	(0.01, 0.30)	(0.32, 0.73)
WNBA	(8.72, 57.19)	(0.00, 0.10)	(0.13, 0.44)	(0.37, 0.89)

**Panel B: Simple Percentile Intervals**

Sport	ROI %	EV Threshold	Epsilon Value	% of Games Bet
NFL	(9.40, 40.60)	(0.00, 0.06)	(0.10, 0.36)	(0.24, 0.81)
NBA	(4.18, 16.87)	(0.00, 0.03)	(0.17, 0.46)	(0.21, 0.56)
NCAAB	(2.07, 16.94)	(0.01, 0.03)	(0.14, 0.37)	(0.29, 0.60)
NCAAF	(1.69, 17.08)	(0.00, 0.08)	(0.01, 0.30)	(0.31, 0.72)
WNBA	(8.73, 47.45)	(0.01, 0.15)	(0.13, 0.32)	(0.35, 0.91)

**Panel C: Weighted High Density Intervals**

Sport	ROI %	EV Threshold	Epsilon Value	% of Games Bet
NFL	(7.04, 35.81)	(0.00, 0.03)	(0.11, 0.36)	(0.38, 0.82)
NBA	(3.74, 16.00)	(0.00, 0.02)	(0.19, 0.46)	(0.28, 0.47)
NCAAB	(1.36, 16.08)	(0.01, 0.03)	(0.14, 0.37)	(0.29, 0.60)
NCAAF	(1.10, 16.00)	(0.00, 0.08)	(0.01, 0.30)	(0.32, 0.73)
WNBA	(8.72, 57.19)	(0.00, 0.10)	(0.13, 0.44)	(0.37, 0.89)

**Panel D: Weighted Percentile Intervals**

Sport	ROI %	EV Threshold	Epsilon Value	% of Games Bet
NFL	(5.92, 41.32)	(0.00, 0.05)	(0.08, 0.34)	(0.30, 0.85)
NBA	(1.66, 25.58)	(0.00, 0.04)	(0.12, 0.49)	(0.11, 0.60)
NCAAB	(2.51, 19.49)	(0.00, 0.03)	(0.14, 0.38)	(0.23, 0.58)
NCAAF	(2.03, 19.68)	(0.00, 0.05)	(0.01, 0.41)	(0.36, 0.73)
WNBA	(6.43, 43.06)	(0.01, 0.15)	(0.13, 0.32)	(0.38, 0.93)

This table displays two-sided 95% confidence intervals for the bootstrapped data. Panel A and B highlight the intervals for bets relying upon the Simple win probability model, and Panel C and D highlight the intervals for bets relying upon the Weighted win probability model. Both Panel A and C utilize 95% percentile intervals, and Panel B and D utilize 95% high density intervals. All ROI and Epsilon Value intervals for all sports in each panel do not contain zero.

Table 6: **One Sided 99% Bootstrap Confidence Intervals****Panel A: Simple Percentile Intervals**

Sport	ROI %	EV Threshold	Epsilon Value	% of Games Bet
NFL	(8.11, 82.85)	(0.00, 0.10)	(0.06, 0.47)	(0.20, 0.84)
NBA	(3.41, 40.25)	(0.00, 0.04)	(0.13, 0.50)	(0.14, 0.62)
NCAAB	(1.20, 33.48)	(0.00, 0.03)	(0.04, 0.47)	(0.25, 0.66)
NCAAF	(0.92, 38.44)	(0.00, 0.08)	(0.01, 0.46)	(0.27, 0.75)
WNBA	(7.69, 145.12)	(0.00, 0.20)	(0.13, 0.50)	(0.30, 0.94)

**Panel B: Weighted Percentile Intervals**

Sport	ROI %	EV Threshold	Epsilon Value	% of Games Bet
NFL	(6.47, 111.64)	(0.00, 0.11)	(0.08, 0.37)	(0.27, 0.88)
NBA	(1.77, 48.90)	(0.00, 0.04)	(0.07, 0.50)	(0.11, 0.62)
NCAAB	(2.63, 35.02)	(0.00, 0.03)	(0.00, 0.46)	(0.27, 0.77)
NCAAF	(2.34, 66.26)	(0.00, 0.08)	(0.01, 0.48)	(0.19, 0.75)
WNBA	(6.80, 99.31)	(0.00, 0.22)	(0.05, 0.50)	(0.33, 0.96)

This table displays one-sided 99% confidence intervals for bootstrapped data. Panel A highlights the intervals for bets relying upon the Simple win probability model, and Panel B highlights the intervals for bets relying upon the Weighted win probability model. All ROI intervals for all sports in each panel do not contain zero. As we are utilizing an one-sided interval, the null and alternative hypotheses for ROI, EV Threshold, and Epsilon Value are respectively:  $H_0 = 0$  and  $H_A > 0$ . By examining the intervals in the table and using the Boneferroni correction (with  $\alpha = 0.05$ ) for each win probability model, we reject the null hypothesis that the optimal returns from this strategy are equivalent to zero for every sport. By again examining the intervals in the table and using the Boneferroni correction (with  $\alpha = 0.05$ ) for each win probability model, we reject the null hypothesis that the optimal Epsilon Value is equivalent to zero for every sport.

---

**Algorithm 1:** Compute Probability: Simple Average

---

**Input:**  $\pi$ , a mapping of spreads to their respective probabilities

**Output:** Simple average of the probabilities of distinct spreads

1 Let  $\sigma$  be the set of all unique spreads.

2 **return**  $\frac{1}{|\sigma|} \sum_{s \in \sigma} \pi_s$ ;

---

---

**Algorithm 2:** Compute Probability: Weighted Average

---

**Input:**  $\phi$ , a mapping of spreads to their respective frequencies

$\pi$ , a mapping of spreads to their respective probabilities

**Output:** Average of the probabilities of all spreads weighted according to the spread frequencies

1 Let  $\sigma$  be the set of all unique spreads.

2 **return**  $\frac{\sum_{s \in \sigma} \pi_s \phi_s}{\sum_{s \in \sigma} \phi_s}$ ;

---

---

**Algorithm 3:** BettingAlgorithm

---

**Input:**  $\bar{\mu}_f$ , the expected value of winnings if the favorite team is bet on  
 $\bar{\mu}_u$ , the expected value of winnings if the underdog team is bet on  
 $\pi_f$ , the probability that the favorite team wins the game  
 $\pi_u$ , the probability that the underdog team wins the game  
 $\epsilon$ , value of the Epsilon hyper-parameter  
 $\tau$ , value of the EV threshold hyper-parameter  
**Output:** -1, if the algorithm decides to bet on neither team  
0, if the algorithm decides to bet on the underdog team OR  
+1, if the algorithm decides to bet on the favorite team

```
1 if  $\neg(\bar{\mu}_f < 0 \wedge \bar{\mu}_u < 0)$  then
2   if  $\pi_f \geq 0.50 + \epsilon$  then
3     if  $\pi_f \geq \pi_u$  then return 1 ;
4     else return 0 ;
5   else
6     if  $\bar{\mu}_f \geq \bar{\mu}_u$  then
7       if  $\bar{\mu}_f > \tau$  then return 1 ;
8     else
9       if  $\bar{\mu}_u > \tau$  then return 0 ;
10 return -1;
```

---

---

**Algorithm 4:** SpreadBetRandomization

---

**Input:**  $\Gamma$ , set of all games

$P_F$ , points scored by favorite in each game

$P_U$ , points scored by underdog in each game

$MS_F$ , minimum spread for favorite in each game

$MS_U$ , minimum spread for underdog in each game

**Output:** Return on Investment (ROI %)

```
1 foreach  $g \in \Gamma$  do
2   if  $P_F + MS_F > P_U$  then
3      $Victor \leftarrow 1$ ;
4   else if  $P_U + MS_U > P_F$  then
5      $Victor \leftarrow 0$ ;
6   else
7      $Victor \leftarrow 2$ ;
8    $Sim_{Victor} \sim U[0, 1]$ ;
9   if  $Sim_{Victor} > 0.5$  then
10     $Sim_{Victor} \leftarrow 1$ ;
11  else
12     $Sim_{Victor} \leftarrow 0$ ;
13  if  $Victor = Sim_{Victor}$  then
14     $Winnings \leftarrow 100/110$ ;
15  else if  $Victor \neq Sim_{Victor} \wedge Victor \neq 2$  then
16     $Winnings \leftarrow -1$ ;
17  else
18     $Winnings \leftarrow 0$ ;
19  $ROI = \frac{100}{|\Gamma|} \sum_{g \in \Gamma} Winnings_g$ ;
20 return  $ROI$ 
```

---

---

**Algorithm 5:** MoneylineBetRandomization

---

**Input:**  $\Gamma$ , set of all games  
 $P_F$ , points scored by favorite in each game  
 $P_U$ , points scored by underdog in each game  
 $PO_F$ , payout for favorite in each game  
 $PO_U$ , payout spread for underdog in each game  
 $\Theta$ , probability of choosing favorite in each simulation  
**Output:** Return on Investment (ROI %)

```
1 foreach  $g \in \Gamma$  do
2   if  $P_F > P_U$  then
3     |  $Victor \leftarrow 1$ ;
4   else if  $P_U > P_F$  then
5     |  $Victor \leftarrow 0$ ;
6   else
7     |  $Victor \leftarrow 2$ ;
8    $Sim_{Victor} \sim U[0, 1]$ ;
9   if  $Sim_{Victor} > \Theta$  then
10    |  $Sim_{Victor} \leftarrow 1$ ;
11  else
12    |  $Sim_{Victor} \leftarrow 0$ ;
13  if  $Victor = Sim_{Victor} \wedge Victor = 1$  then
14    |  $Winnings \leftarrow PO_F$ ;
15  else if  $Victor = Sim_{Victor} \wedge Victor = 0$  then
16    |  $Winnings \leftarrow PO_U$ ;
17  else if  $Victor \neq Sim_{Victor} \wedge Victor \neq 2$  then
18    |  $Winnings \leftarrow -1$ ;
19  else
20    |  $Winnings \leftarrow 0$ ;
21 ROI =  $\frac{100}{|\Gamma|} \sum_{g \in \Gamma} Winnings_g$ ;
22 return ROI
```

---



---

**Algorithm 6:** BootstrapAlgorithm

---

**Input:**  $\Gamma$ , set of all games

$N_G$ , number of sample games

$EP$ , vector of (0, 95<sup>th</sup> Quantile) of Epsilon Values from bootstrap sample, step size of 0.01

$EV$ , vector of (0, 95<sup>th</sup> Quantile) of EV Thresholds from bootstrap sample, step size of 0.001

**Output:**  $O_{ROI}$ , Optimal Return on Investment (ROI %),

$O_{EP}$ , Optimal Epsilon Value,

$O_{EV}$ , Optimal Expected Value Threshold

```
1  $N_{BS} \leftarrow 0$ ;
2  $\Gamma^{BS} \leftarrow []$ ;
3 while  $N_{BS} \leq N_G$  do
4   randomly select a game,  $g$  from  $\Gamma$ ;
5    $\Gamma^{BS}.append(g)$ ;
6    $N_{BS} \leftarrow N_{BS} + 1$ ;
7 initialize  $TRBS$  as a matrix of zeros with dimensions ( $|EP|$ ,  $|EV|$ );
8 initialize  $ROIBS$  as a matrix of zeros with dimensions ( $|EP|$ ,  $|EV|$ );
9  $i, j \leftarrow 0, 0$ ;
10 while  $i \leq |EP|$  do
11   while  $j \leq |EV|$  do
12      $N \leftarrow 0$ ;
13      $TR \leftarrow 0$ ;
14     foreach  $g \in \Gamma^{BS}$  do
15        $D \leftarrow BettingAlgorithm(g.\bar{\mu}_f, g.\bar{\mu}_u, g.\pi_f, g.\pi_u, EP_i, EV_j)$ ;
16       if  $D \neq 0$  then  $N \leftarrow N + 1$ ;
17       if  $D = -1$  then
18         if  $g.victor = \text{Favorite}$  then  $TR \leftarrow TR - 1$ ;
19         else  $TR \leftarrow TR + g.underdog\_payout$ ;
20       else if  $D = 1$  then
21         if  $g.victor = \text{Favorite}$  then  $TR \leftarrow TR + g.favorite\_payout$ ;
22         else  $TR \leftarrow TR - 1$ ;
23      $TRBS[i, j] \leftarrow TR$ ;
24      $ROIBS[i, j] \leftarrow \frac{TR}{N}$ ;
25      $j \leftarrow j + 1$ 
26    $i \leftarrow i + 1$ 
27  $i', j' \leftarrow \arg \max_{0 \leq i \leq |EP|-1, 0 \leq j \leq |EV|-1} TRBS[i, j]$ ;
28  $O_{ROI} \leftarrow ROIBS[i', j']$ ;
29  $O_{EP}, O_{EV} \leftarrow EP[i'], EV[j']$ ;
30 return  $O_{ROI}, O_{EP}, O_{EV}$ ;
```

---

## ONLINE SUPPLEMENT

# Beating the House: Identifying Inefficiencies in Sports Betting Markets

October 22, 2019

### **Abstract**

We utilize histograms to visualize features of the bootstrap iterations. We also provide the methodology for these visualizations.

# Graph Methodology

## Univariate Histogram Methodology

Keeping with the theme of using non-parametric techniques, we utilized Scott's normal reference rule<sup>1</sup> to determine optimal bin sizing to generate these univariate histograms. These visualizations were very useful to us when we needed to decide which kind of bootstrap intervals we wanted to apply to the bootstrapped data.

## Bivariate Histogram Methodology

As we wish to utilize non-parametric techniques, we chose the bivariate average shifted histogram<sup>2</sup> in order to create 3-D density estimation visualizations of various combinations of variables from the bootstrapped data.

---

<sup>1</sup> Scott, David W. "On optimal and data-based histograms." *Biometrika*, Volume 66, Issue 3, December 1979, Pages 605–610, <https://doi.org/10.1093/biomet/66.3.605>

<sup>2</sup> Scott, David W. "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions". *Ann. Statist.* 13 (1985), no. 3, 1024–1040. doi:10.1214/aos/1176349654

# Univariate Histograms

Figure 1: Simple NFL Histograms

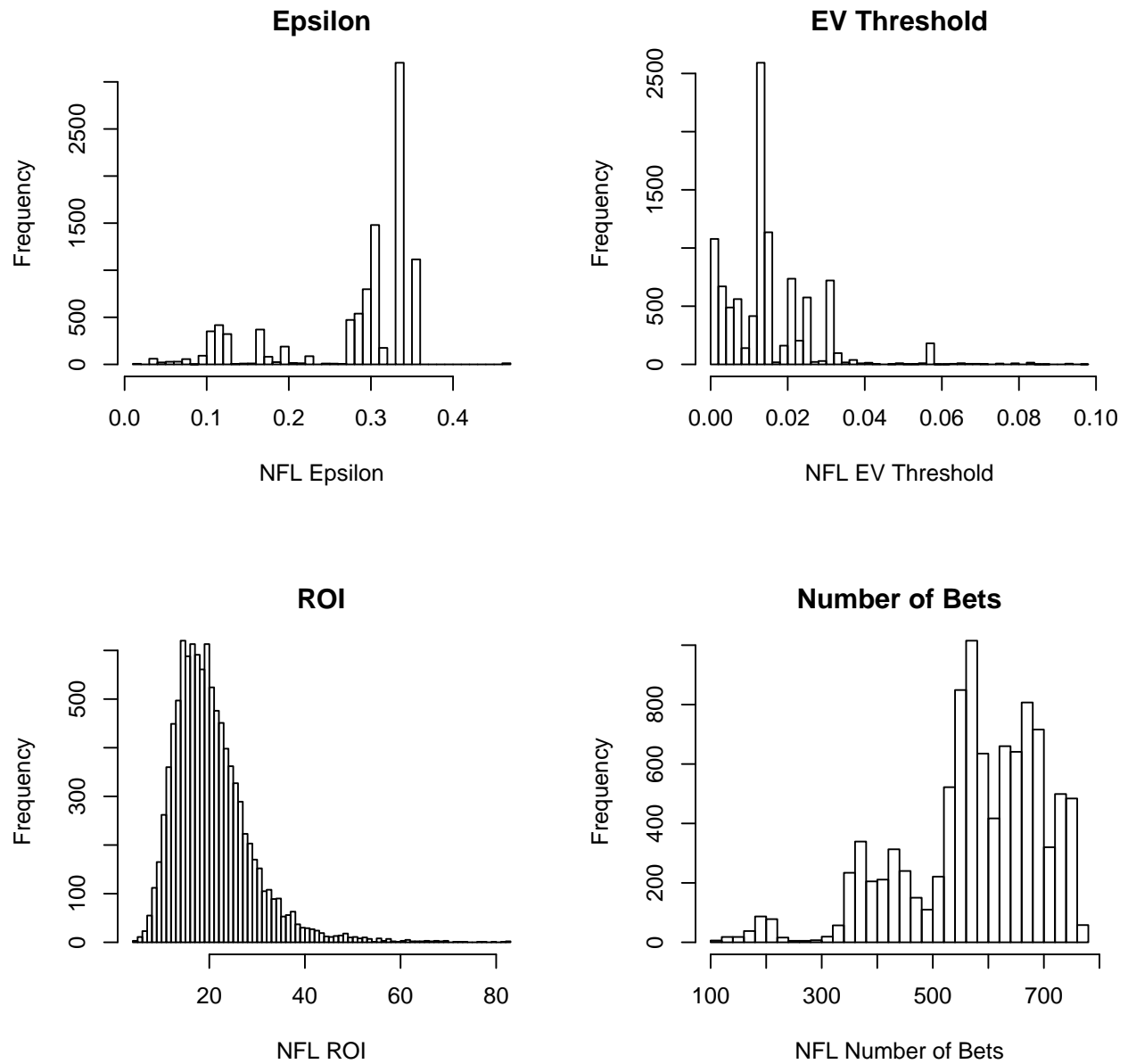


Figure 2: **Weighted NFL Histograms**

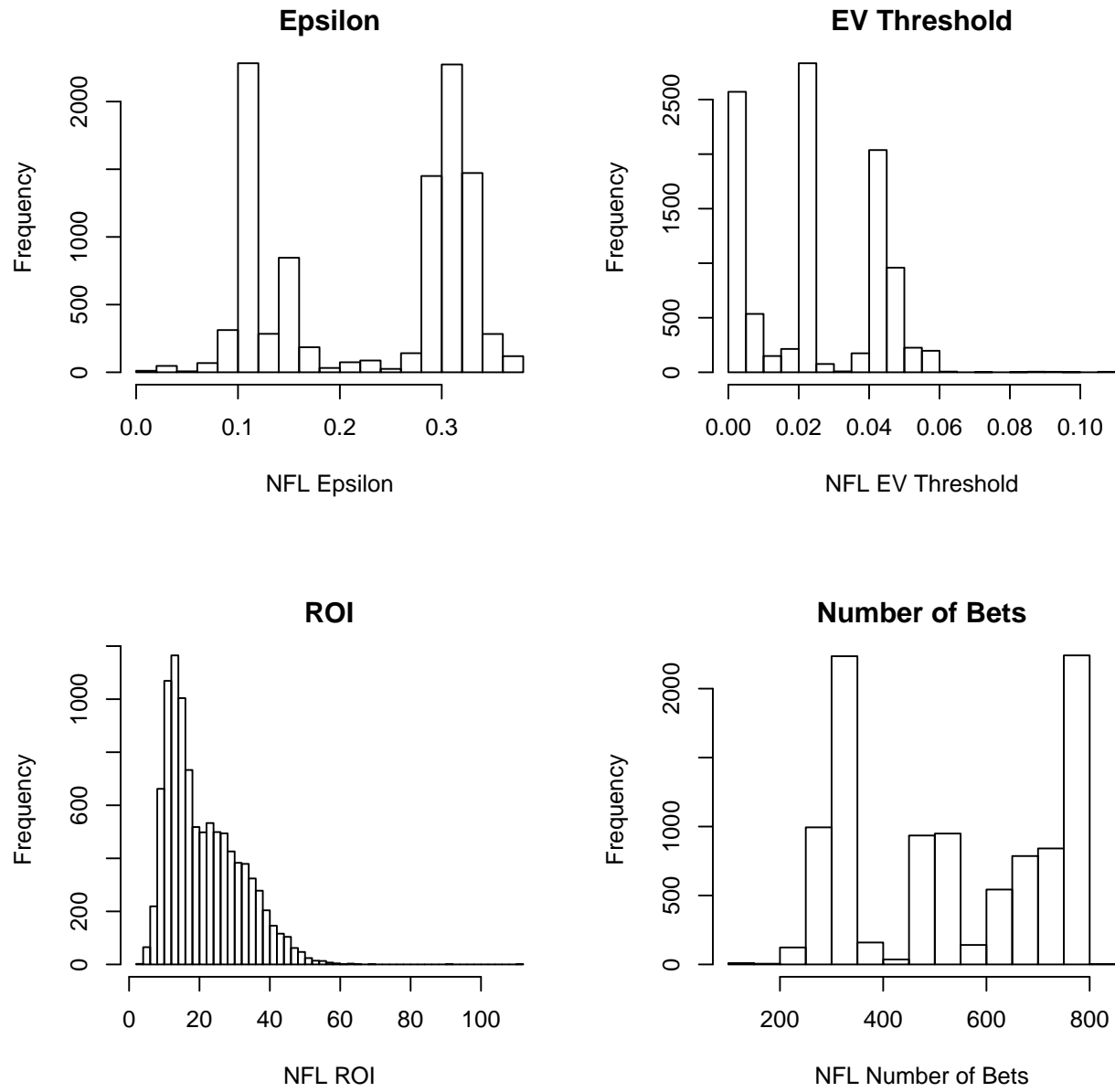


Figure 3: Simple NBA Histograms

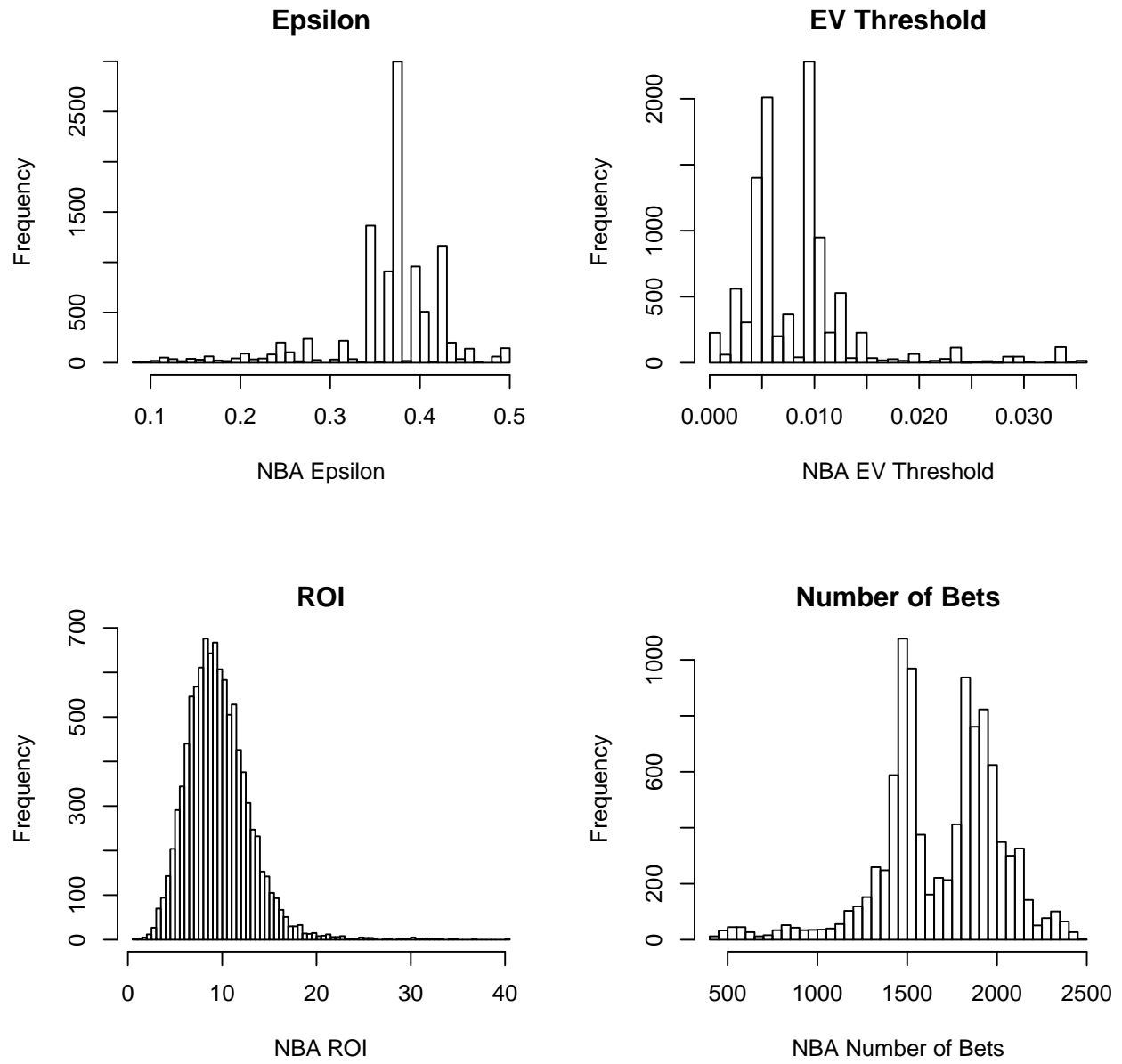


Figure 4: **Weighted NBA Histograms**

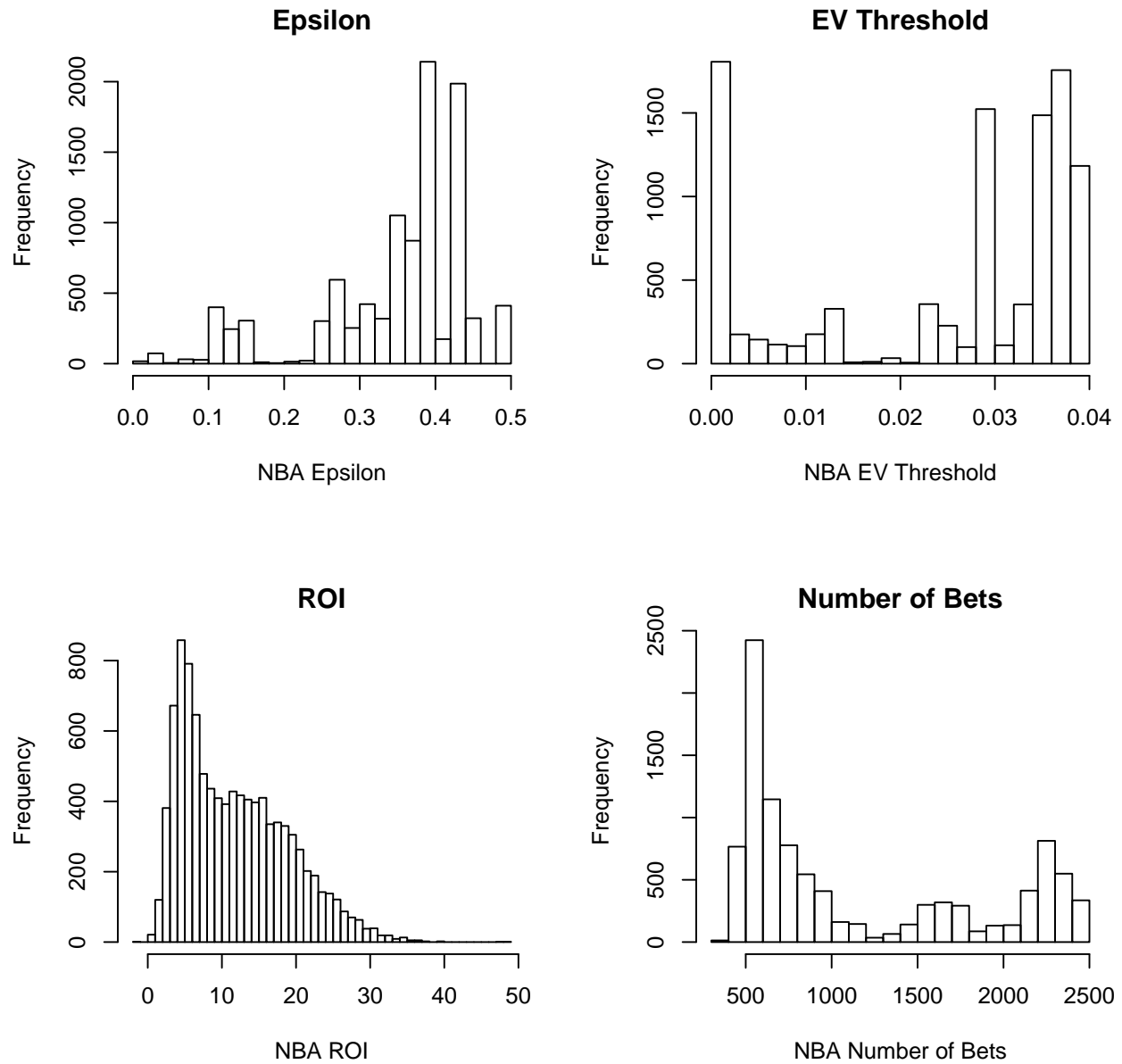


Figure 5: Simple NCAAB Histograms

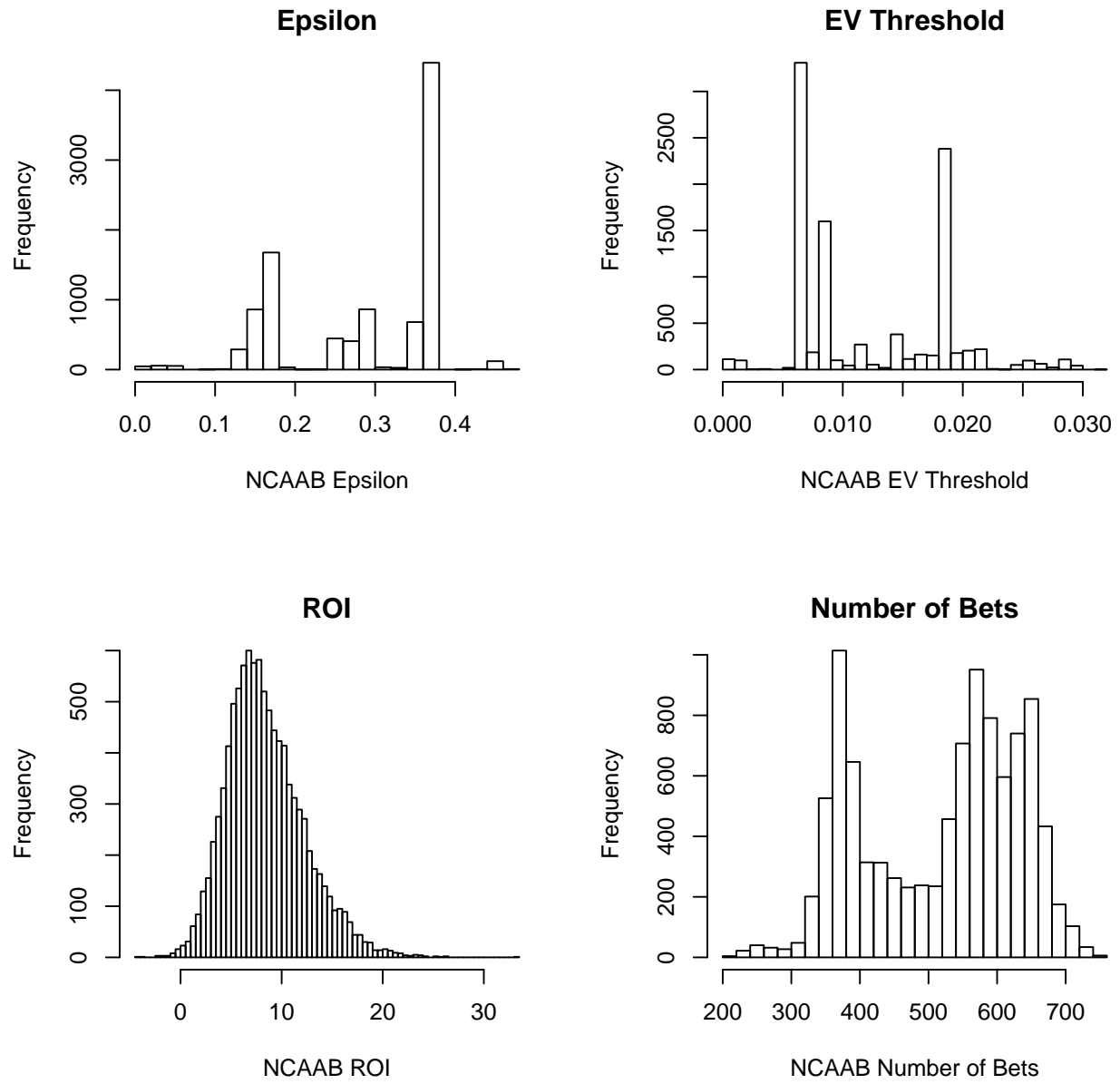




Figure 6: **Weighted NCAAB Histograms**

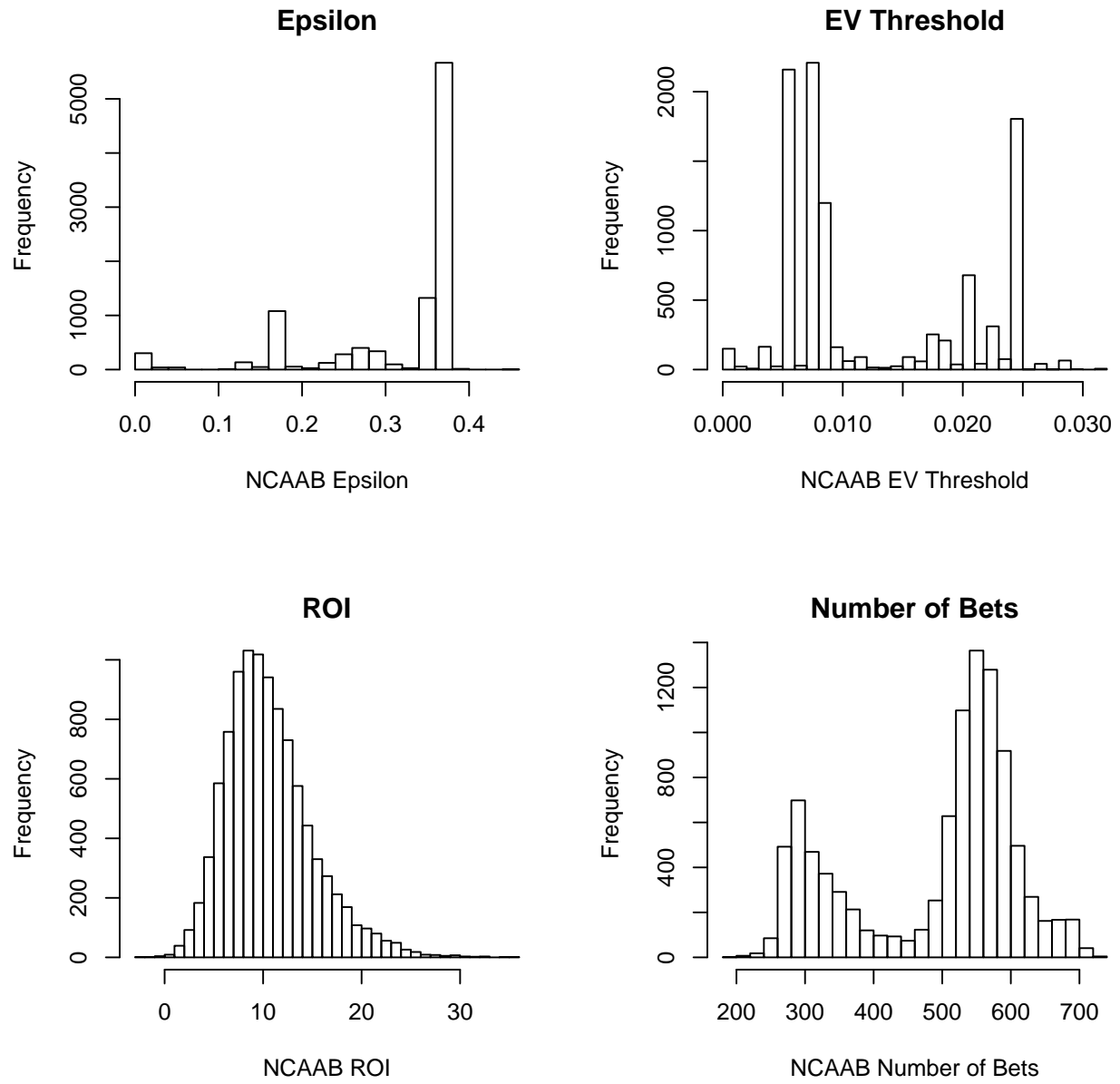


Figure 7: Simple NCAAF Histograms

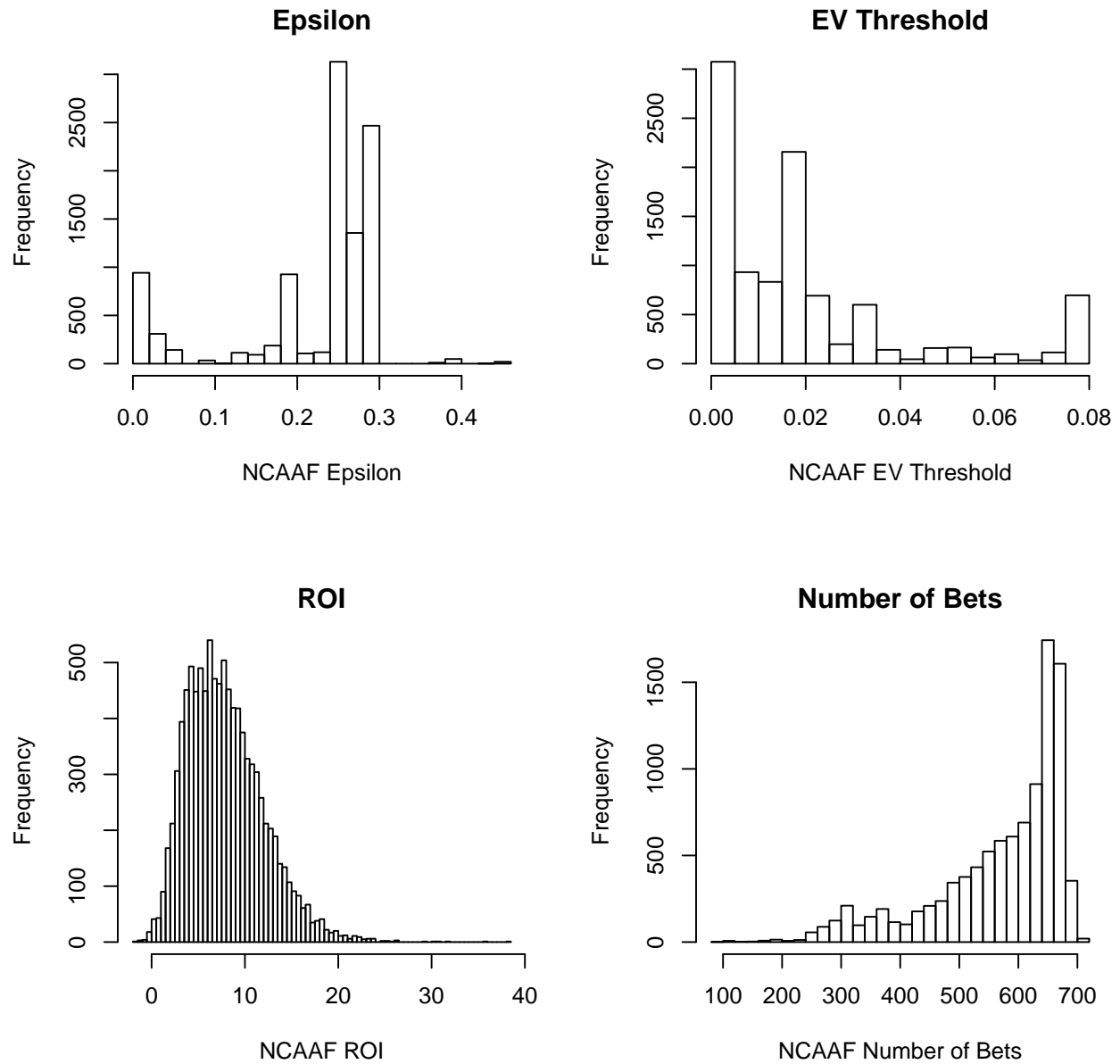


Figure 8: **Weighted NCAAF Histograms**

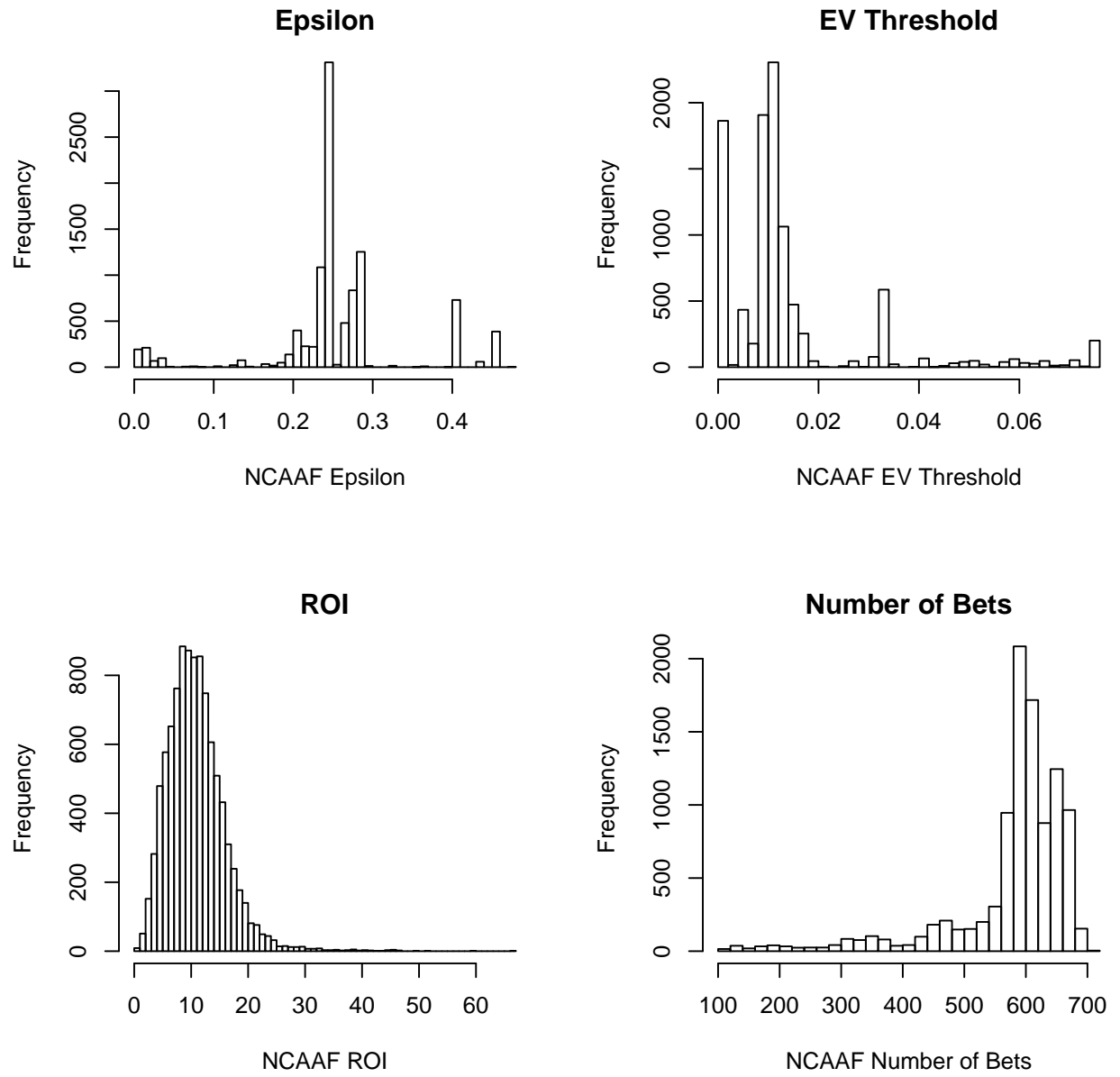


Figure 9: Simple WNBA Histograms

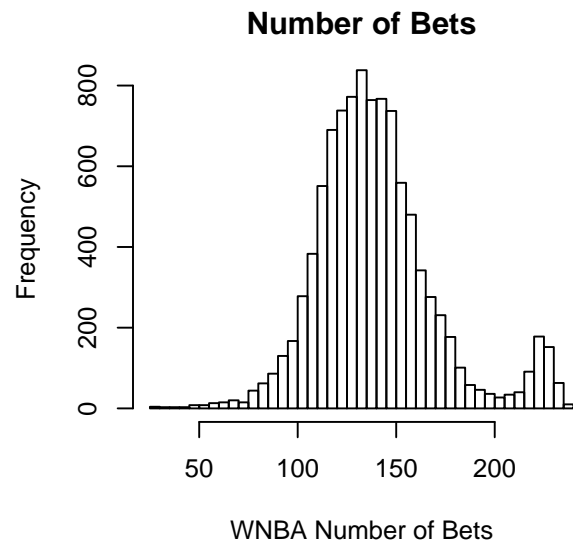
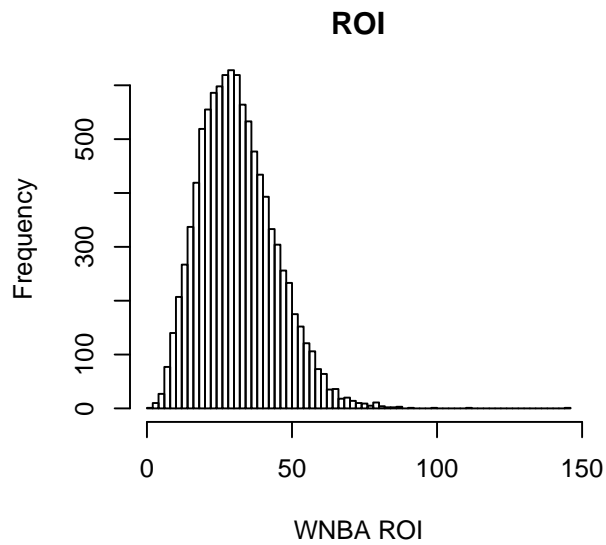
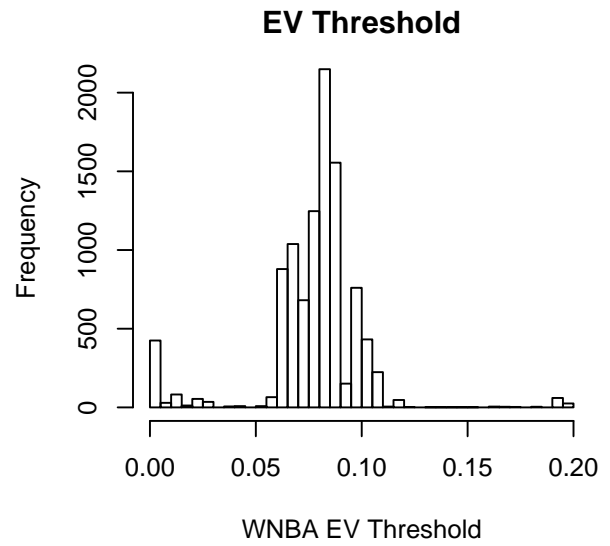
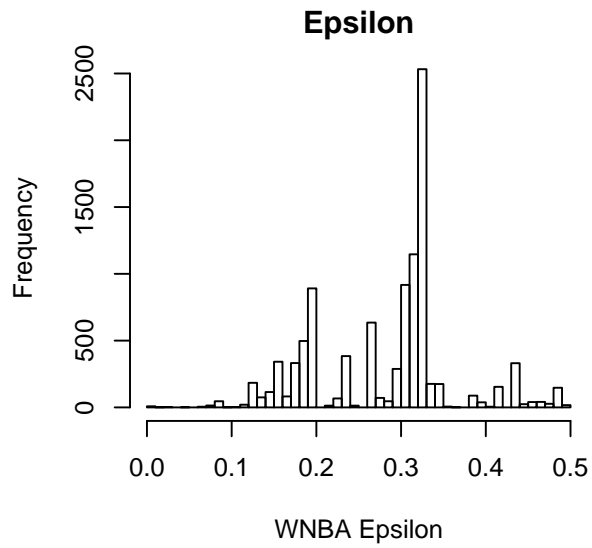
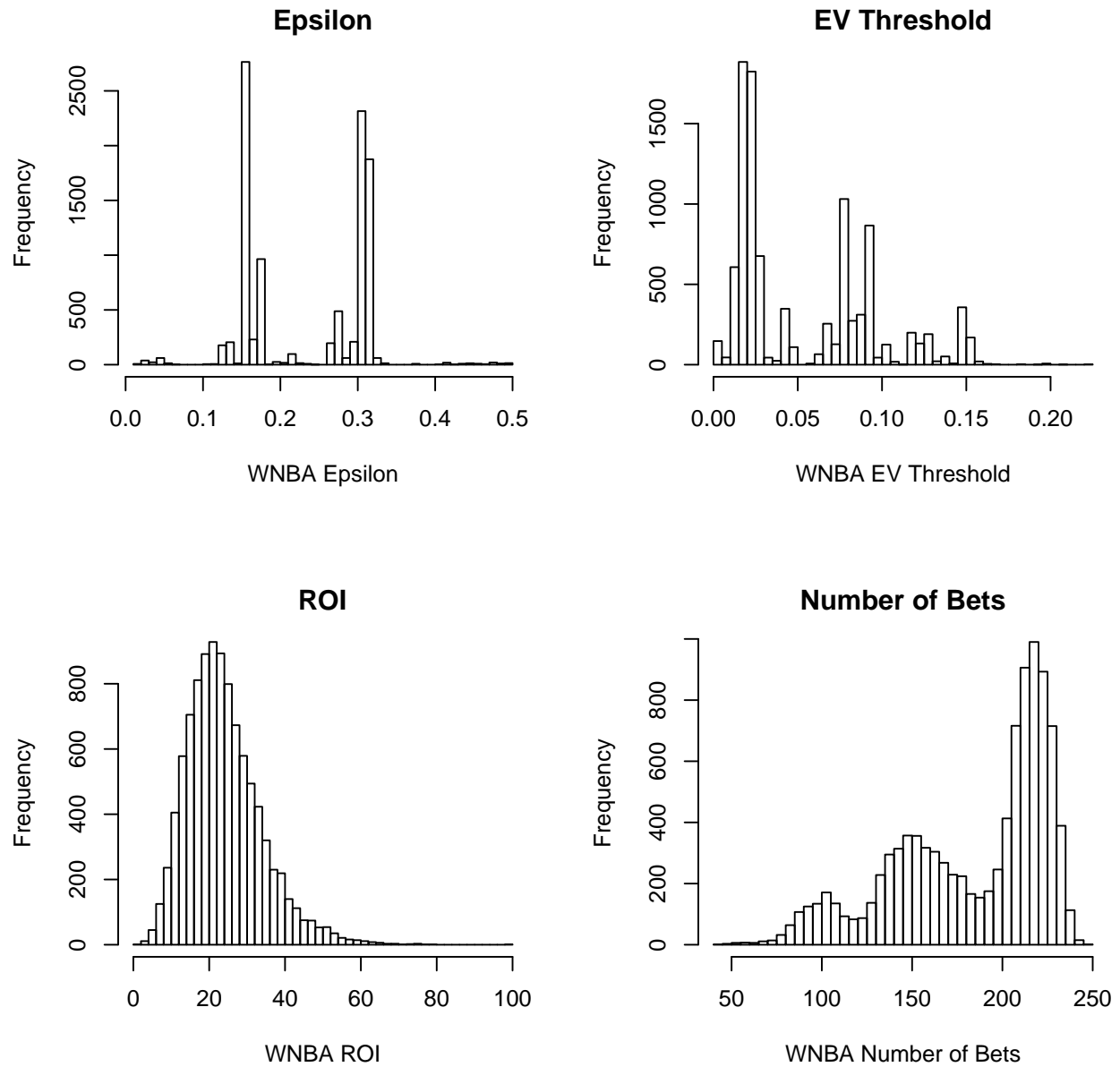


Figure 10: **Weighted WNBA Histograms**



# Bivariate Histograms

Figure 11: Simple NFL Histograms

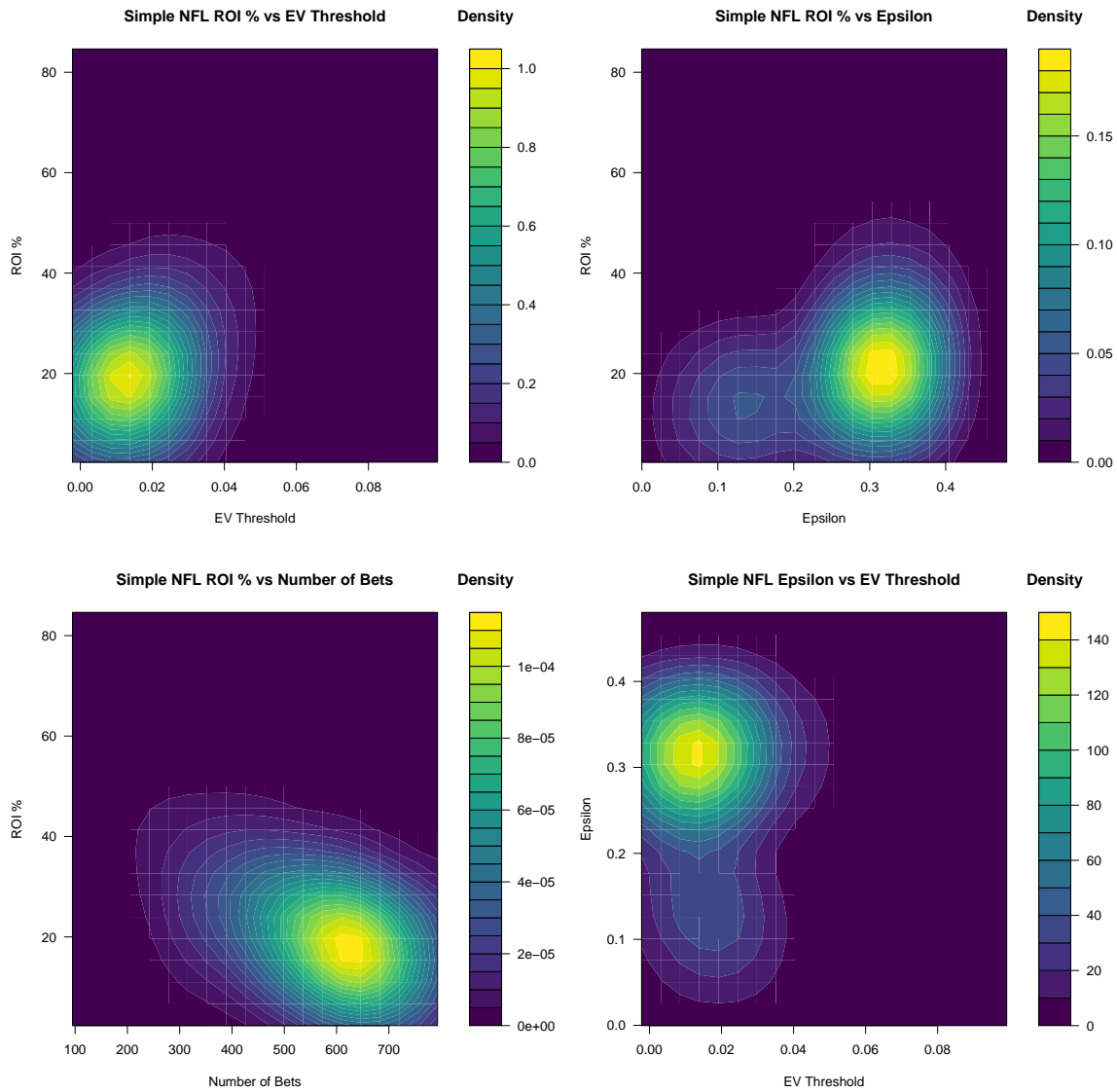


Figure 12: Weighted NFL Histograms

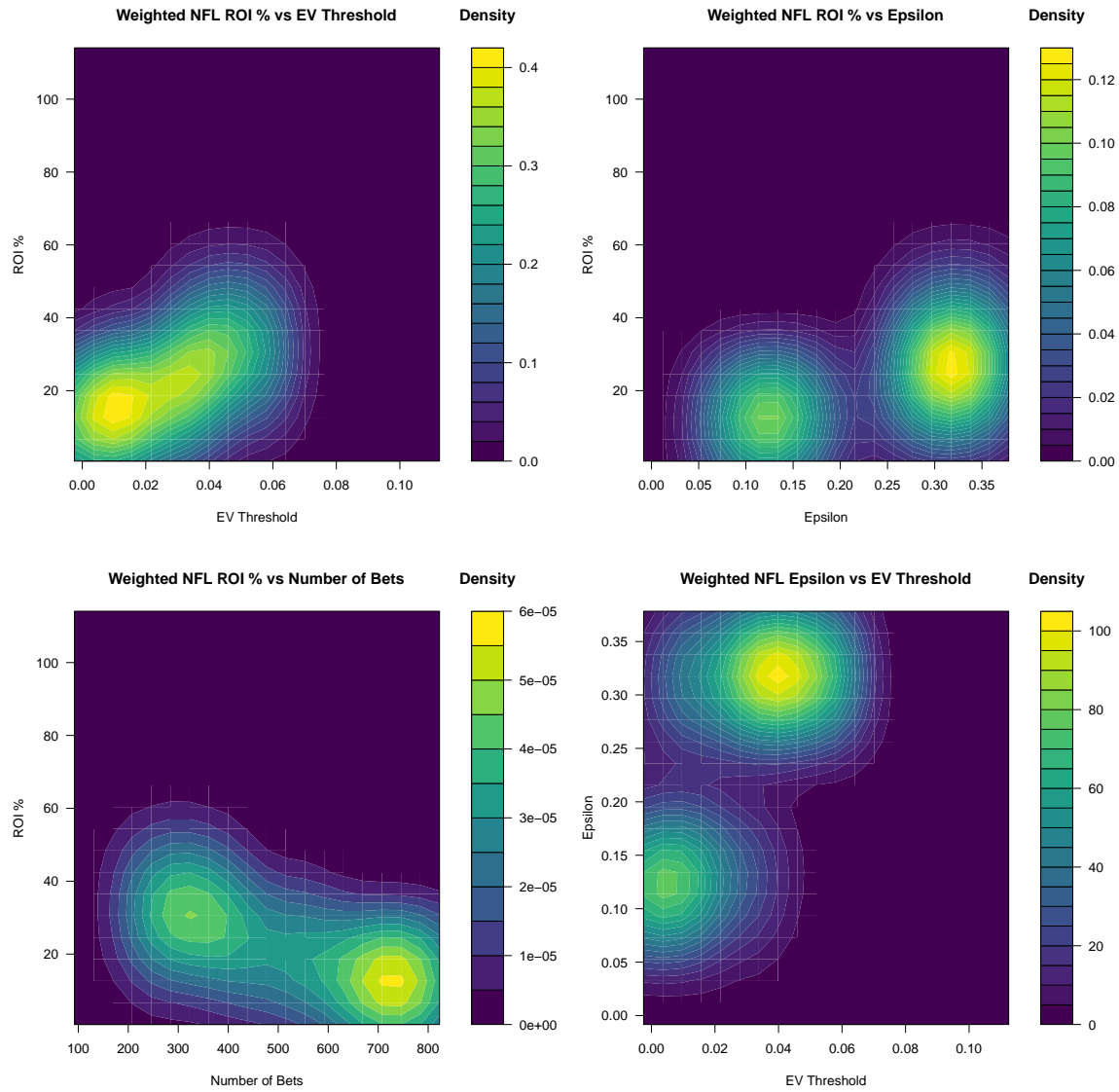


Figure 13: Simple NBA Histograms

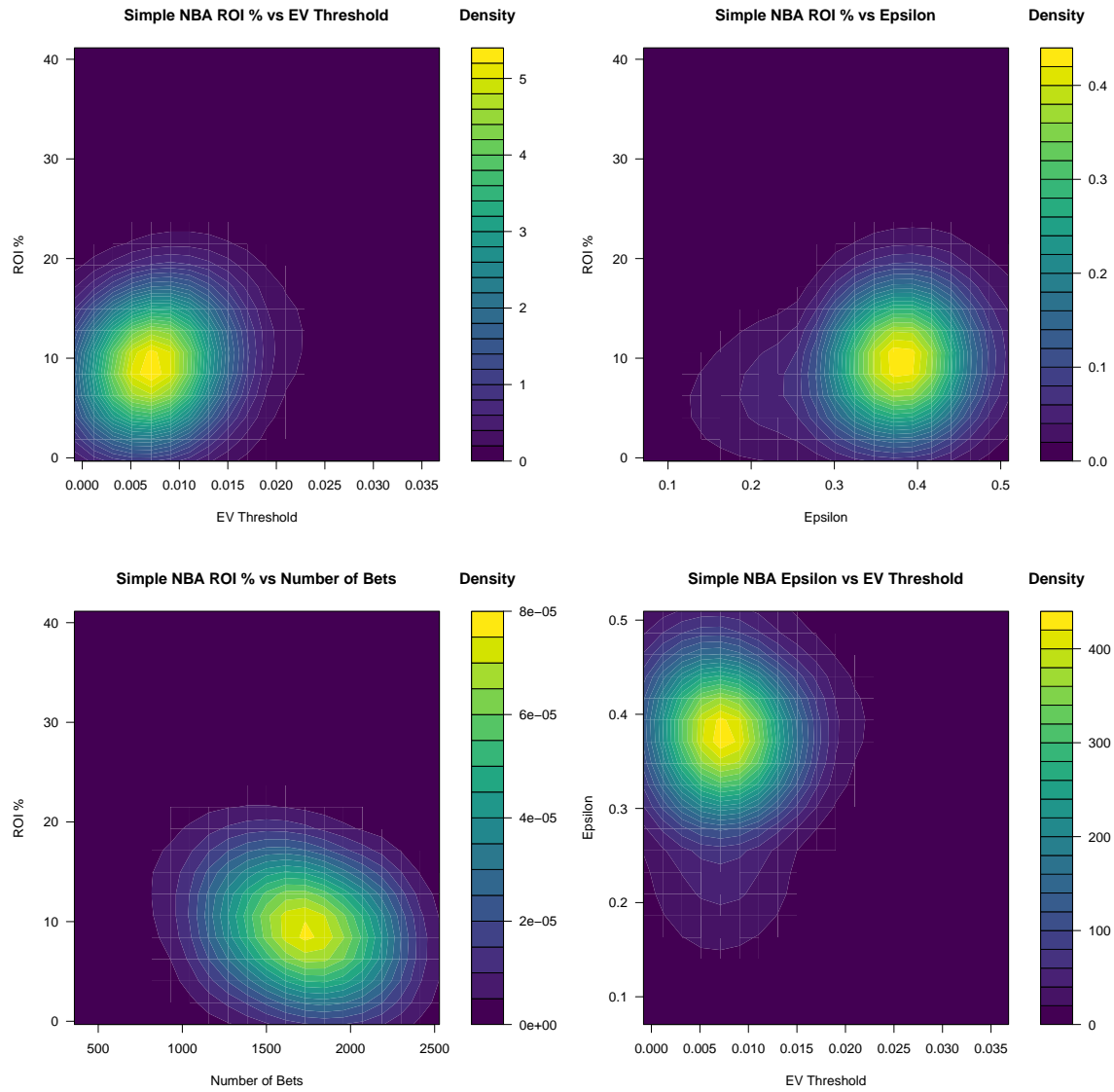




Figure 14: Weighted NBA 2-D Histograms

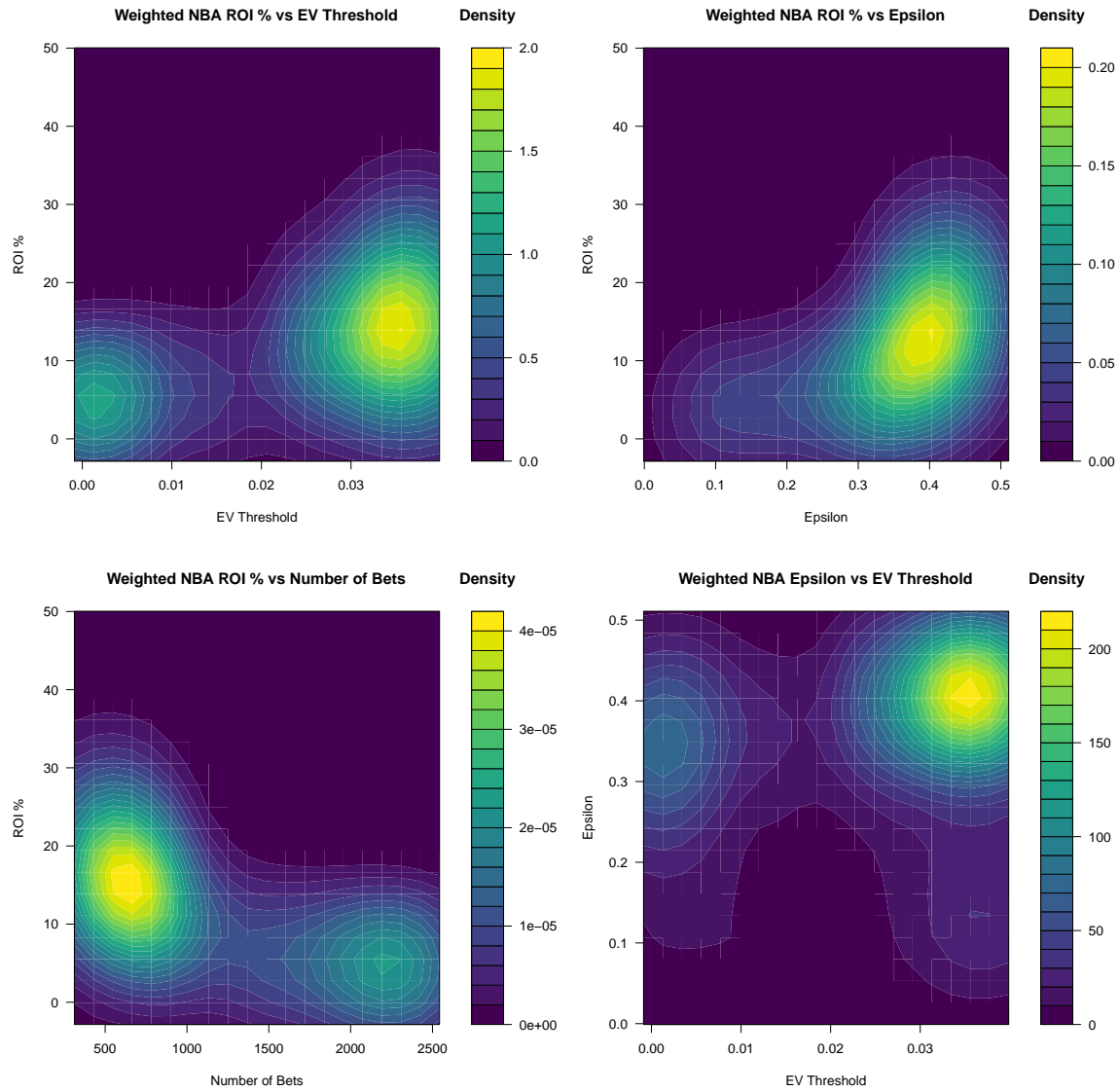


Figure 15: Simple NCAAAB Histograms

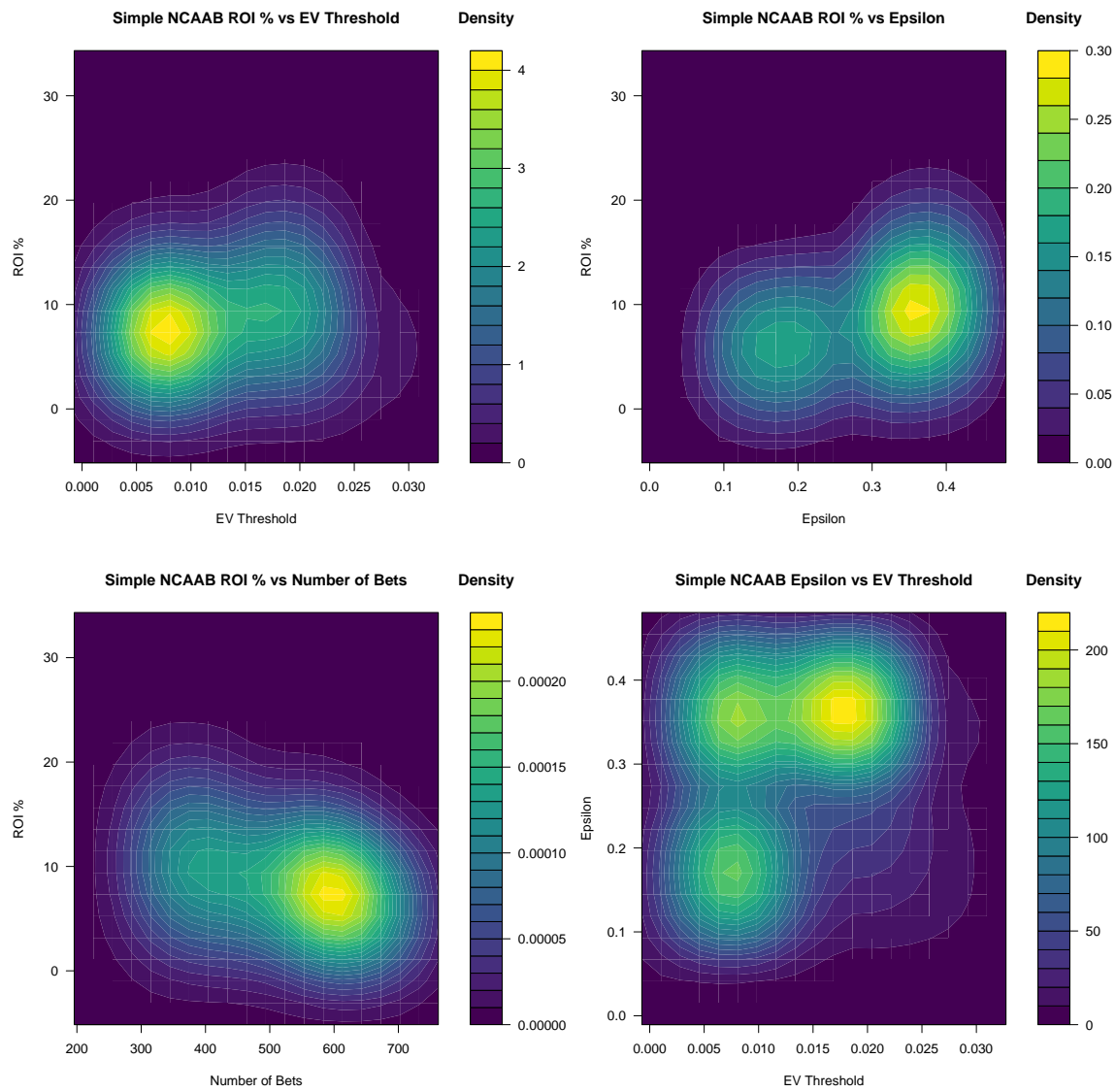


Figure 16: Weighted NCAAB Histograms

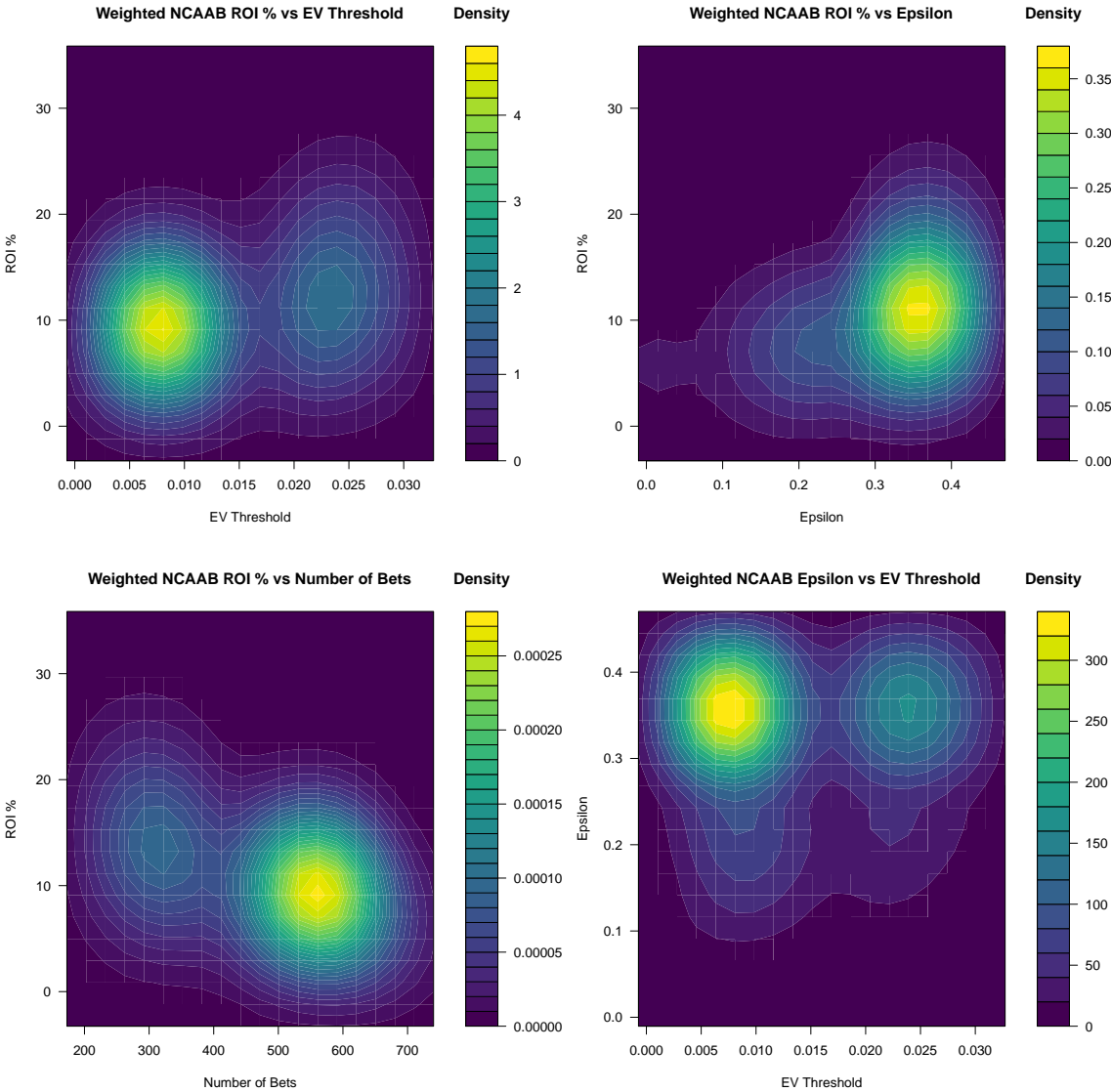


Figure 17: Simple NCAAF Histograms

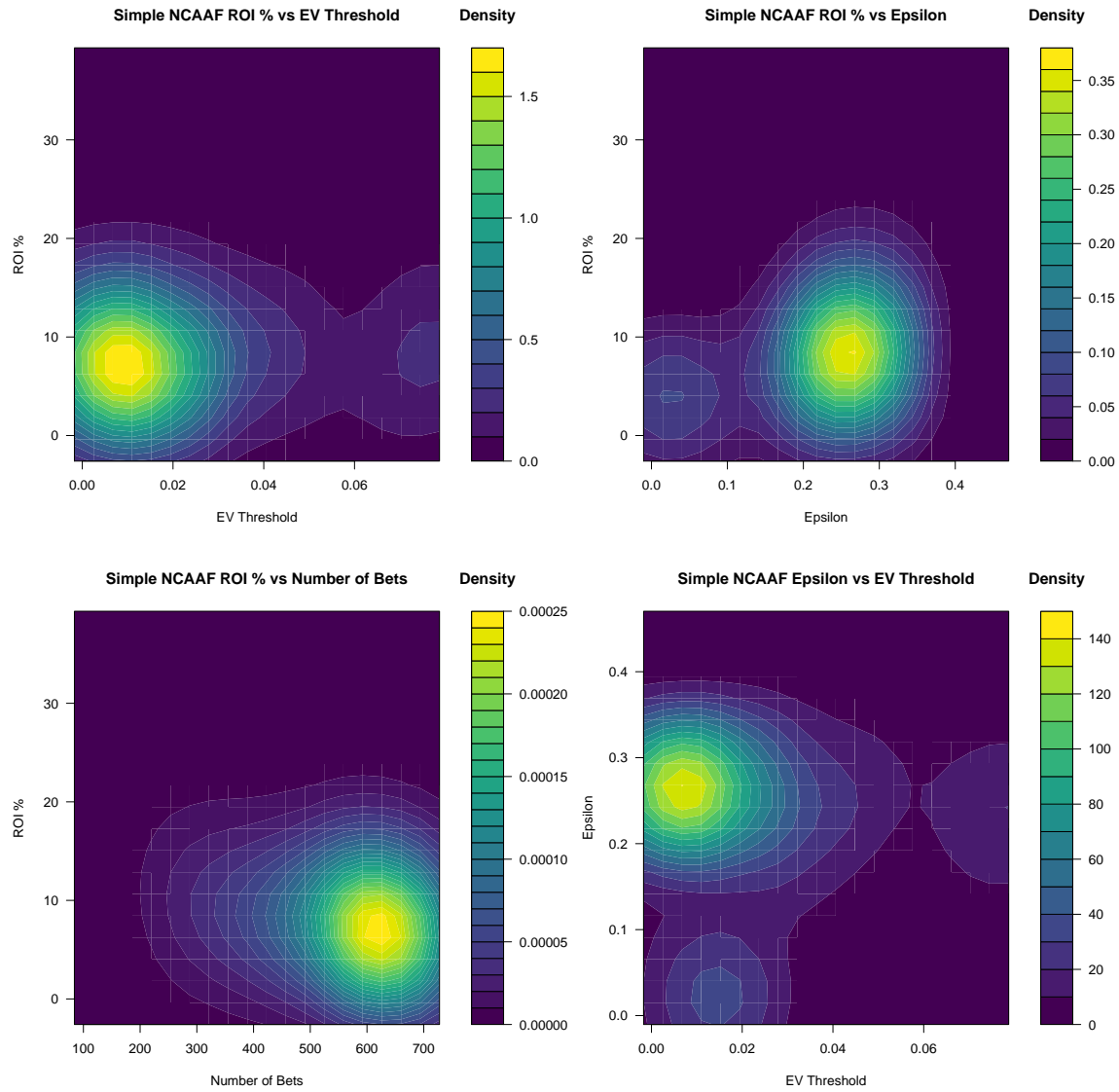


Figure 18: Weighted NCAAAF Histograms

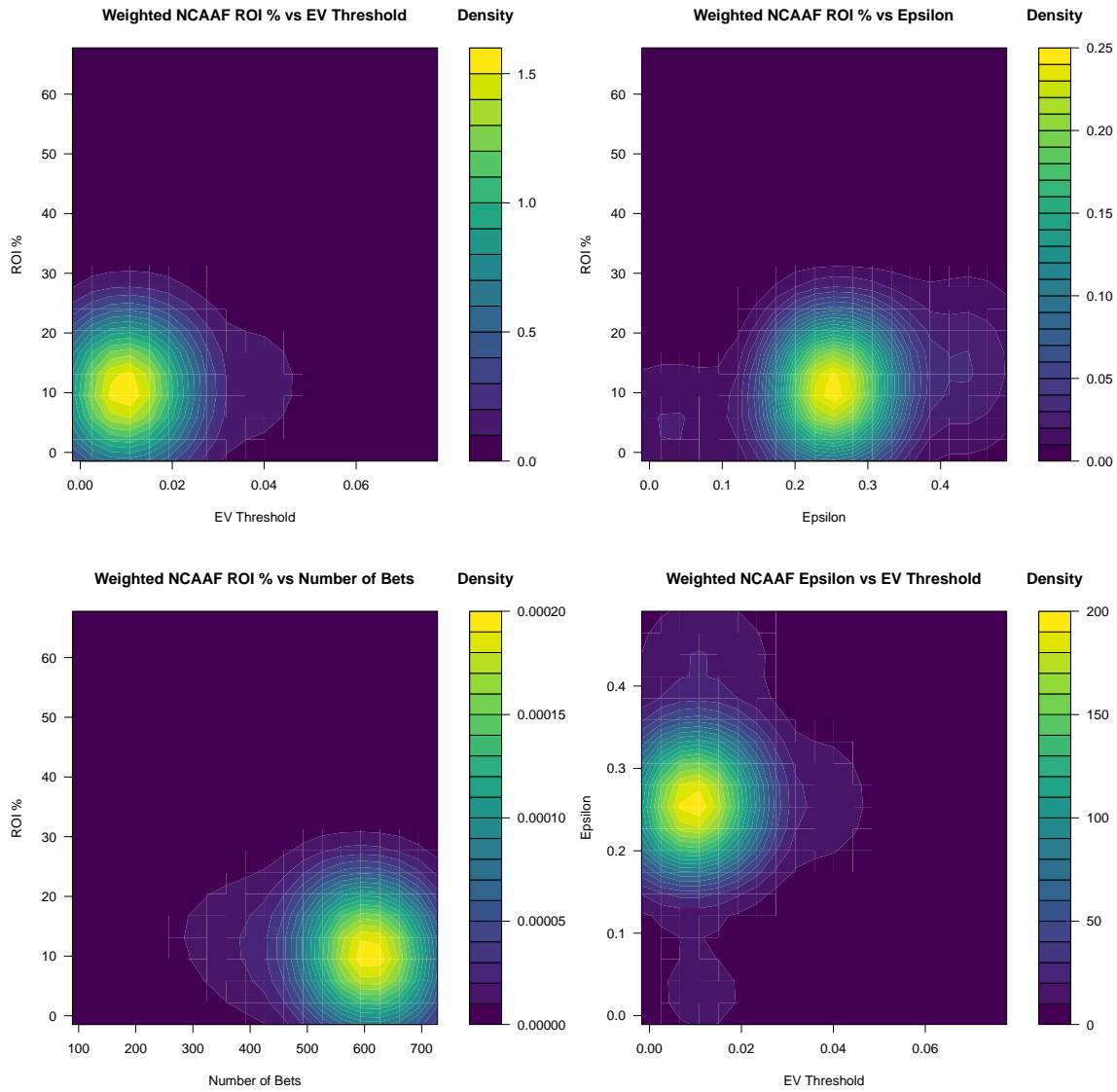


Figure 19: Simple WNBA Histograms

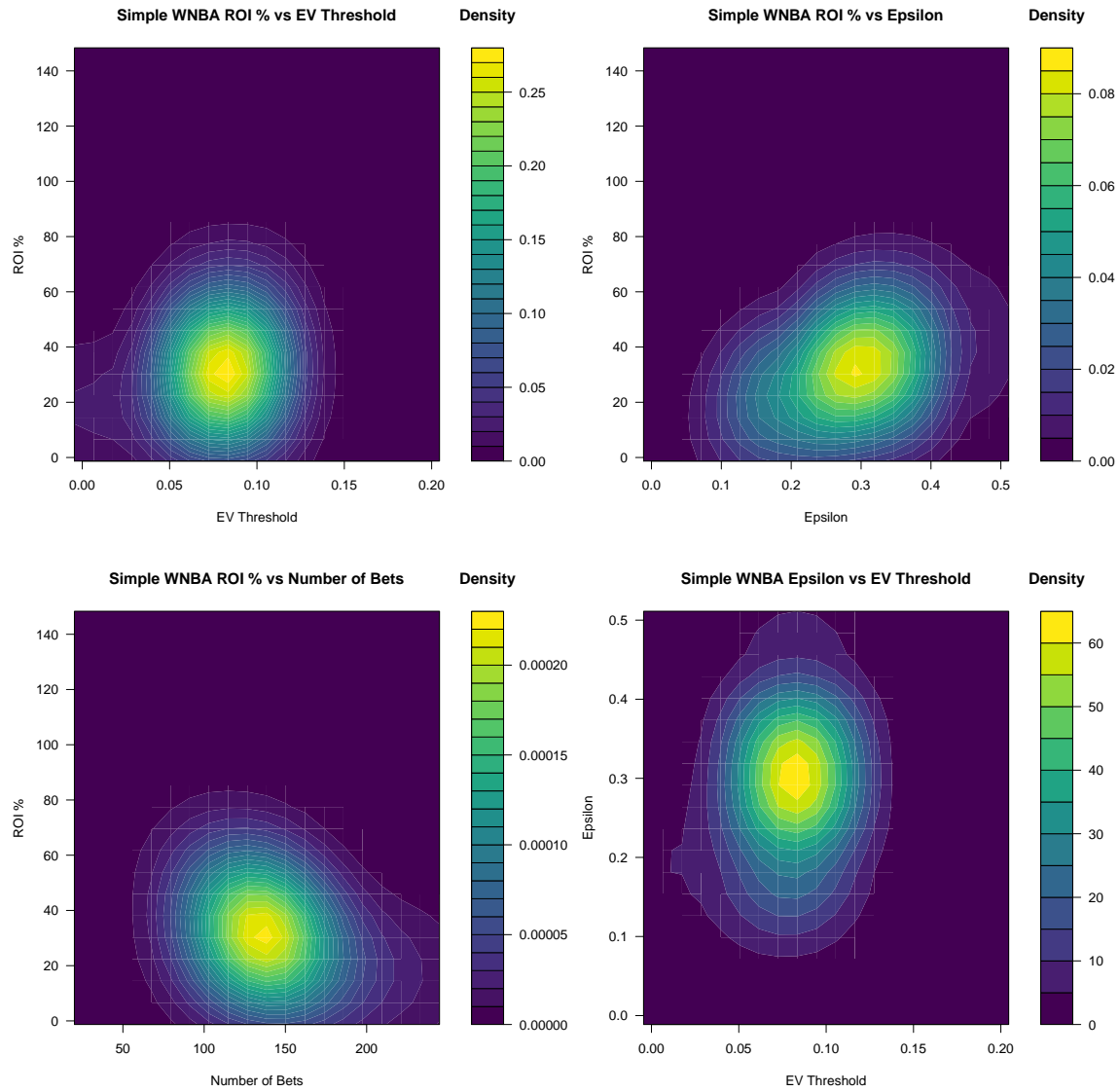


Figure 20: Weighted WNBA Histograms

