On the Efficiency of Racetrack Betting Market: A New Test for the Favorite-Longshot Bias

Jinook Jeong[*] · Jee Young Kim[†] · Yoon Jae Ro[°]

Abstract

A number of empirical researches on the efficiency of racetrack betting market have shown the 'favorite-longshot bias,' which means longshots are overbet while favorites are underbet. Asian markets such as Hong Kong and Japan, however, have produced some contradictory empirical evidence to the bias. One critical element in the efficiency test procedure is how to assess the unobservable objective winning probability of a horse in a race. This paper proposes a new test framework with a more general evaluation of the objective probability of winning than the traditional method. Unlike the traditional method, our model allows the heterogeneity of the horses and the races. We apply the new empirical method to test whether the favorite-longshot bias is present in racetrack betting market of Korea. We found that the favorite-longshot bias exists in the racetrack market of Korea and the result distinguishes Korean racetrack market from other Asian markets.

Keywords: Market efficiency test, Favorite-longshot bias, Racetrack betting, Probit regression

JEL Classification: G25, D81, G14, L83

Acknowledgements: This study was partially funded by the National Research Foundation of Korea (Grant Number NRF-2013S1A3A2053586/NRF-2016S1A3A2923769). The authors declare that they have no conflict of interest.

[*] School of Economics, Yonsei University, e-mail: jinookjeong@gmail.com

[†] School of Economics, Yonsei University, e-mail: jyk33@georgetown.edu

[°] Department of Economics, University of California at Riverside, e-mail: yoon.ro@email.ucr.edu

1 Introduction

The sport betting market is intriguing since it combines two major human interests: gambling and sports. Spectators watch the game and try to predict the wining team by past observation on the ability of the players, and this makes sport betting different from the lottery or casino betting. In sport gambling, the payoff distribution is observable: the bettors observe the market odds while making their betting decision. That is why the sport betting market has attracted a lot of attention from various fields.

Testing the market efficiency has been a major research topic in the sport betting market. If the market is efficient, the average returns to the bettors should be equal. In reality, however, many researchers have empirically found higher expected returns at lower odds than at higher odds (Griffith 1949).[1] A majority of sport market efficiency studies have focused on racetrack markets due to the advantages of racetrack data: (1) each bet has a well-defined termination point at which its value becomes certain; (2) the conditions have a better chance of being efficient because of the quick and repeated feedback which tends to facilitate learning (see Thaler and Ziemba 1988 for details); and (3) the public availability of the data.

Many economists have empirically tested the market efficiency hypothesis using the racetrack market, and most of them have concluded that the market is inefficient. Of many anomalies in the racetrack market, the dominant theme is known as 'favorite-longshot' bias (hereafter, FLB) wherein the bettors consistently overbet longshots and underbet favorites. A

[1] The studies of market efficiency in sport betting markets include basketball (Camerer, 1989; Brown and Sauer ,1993; Paul and Weinbach, 2005), baseball(Woodland and Woodland, 1994), football (Pope and Peel,1989; Golec and Tamarkin, 1991; Forrest and Simmons, 2000; Cain et al., 2000, 2003; Forrest, Goddard, and Simmons, 2005; Deschamps and Gergaud, 2012; Graham and Stott, 2008; Vlastakis et al., 2009; Borghesi, 2012; Koning, 2012; Direr, 2013; and Nyberg, 2014), ice hockey (Gandar et al.,1988; Dare and MacDonald,1996; Gray and Gray,1997; Woodland and Woodland, 2001; Gandar et al., 2004; Paul and Weinbach, 2012) and tennis (Cain et al., 2003; Lahvička, 2014; and Abinzano et al., 2017) among others.

number of studies have documented the existence of FLB in racetrack markets around the world.[2]

While the existence of FLB is well supported for the western racetrack betting market, some Asian countries exhibit a notable exception to these findings. For example, the reversed bias is detected for Hong Kong and Japanese racetrack betting markets (Busche and Hall 1988; Busche 1994; Walls and Busche 2003). In Hong Kong and Japan, the favorites are being overbet rather than underbet. This reversed bias from Hong Kong and Japan is still an unsolved puzzle with little theoretical analysis.[3]

This paper aims to contribute to the FLB literature in two aspects. First, we propose a new methodology to estimate the unobservable objective probability of winning. The previous studies employ a proxy for the unobservable objective probability of winning after Ali (1977). We suggest a more general and stable method to estimate the unobservable probability of winning. Second, we use previously unexplored Korean racetrack market data. No empirical study of FLB in Korean racetrack betting market has been conducted, except two unpublished manuscripts.[4] Probably the short history of racetrack betting and limited data availability are the reasons why Korean market has not been investigated. As Korea has a similar culture to Hong Kong and Japan where the reverse FLB has been detected, we expect Korean data will provide a clue to the puzzle, especially on whether the reverse FLB is a cultural phenomenon.

The remainder of this paper is organized as follows: section 2 reviews the background of the study and the related literature; section 3 provides an overview of the Korean racetrack betting market and outlines the data employed in the analysis; section 4 illustrates our empirical

---

[2] FLB has been found in the US (Ali, 1977; Gramm, 2005), Australia (Tuckwell, 1983), Canada (Hausch et al., 1981), Germany (Winter and Kukuk, 2006), New Zealand (Gandar et al., 2001), and the UK (Williams and Paton, 1997; Bruce and Johnson 2000).

[3] There are studies on reverse FLB in other sport betting markets outside of horse tracks, including Swidler and Shaw (1995) and Schnytzer and Weinberg (2008).

[4] Those two manuscripts are Kim (2008) and Ro (2014).

model used for the market efficiency test and presents the estimation results; finally, in Section 5 we summarize our conclusions.

2. Theoretical Background and the Previous Studies

2.1. The Favorite-Longshot Bias

Market efficiency hypothesis of betting market is based on the study of Fama (1970). According to Fama's hypothesis, financial markets are efficient. The efficient markets are perfectly competitive because the prices in the markets are assumed to follow a random walk and completely reflect all the available information. Under the hypothesis, theoretically, it is useless to predict future prices. If all the participants were completely rational and perfect information is available to them, the future prices could not be estimated and no one can consistently earn returns above average on a risk adjusted basis.[5] Many researchers explain that betting market has similar features of the stock market and better suites to test market efficiency (Williams, 1999; Thaler and Ziemba, 1998).

Following Fama (1970), the efficient market hypothesis in racetrack betting market explains the expected return on any bet would be the same regardless of the horse on which one bets. All participants are so rational that they bet on the horse with the highest probability of winning. If many people bet on the horse with higher probability to winning, the actual return of that horse should be lower than the odds of the other horses.

However, many studies have found counter examples known as the FLB, which describes that betting on the horse with the highest probability of winning actually yields a higher return. Griffith (1949) first reported that the realized average rates of returns from

[5] Even though the market efficiency hypothesis allows market prices to be imperfect in the short term, there exists the literature showing the true values will win out in the long term. (Malkiel, 2003; Gray et al. 2005)

betting on favorite horses tend to be robustly and significantly greater than those from betting on long shot horses in American horse races.

Existing explanations associated with FLB can be divided into two categories: demand side explanations and supply side explanations. The demand side reasons can be further classified two explanations: Bettor's risk preferences and behavioral economics under asymmetric information. First, a number of studies in the literature concern risk preference in a representative agent model with risk loving utility functions (Weitzman, 1965; Ali, 1977; Quandt, 1986; Kanto et al., 1992; Bruce and Johnson, 1992; Golec and Tamarkin, 1998; Jullien and Salane, 2000).[6] Risk preferences of players, described by the risk preference model, are determined by the relationship between the objective probability of winning a race and the subjective probability reflected in the track odds. The other stream, strategic behavior among bettors exemplified in the information model, explains the misperceived probabilities associated with different outcomes and examines the market efficiency of racetrack (Shin, 1991; Paton, 1998; Ternell and Farmer, 1996; Hurley and McDonough, 1995; and Ottaviani and Sorensen 2010).

On the other hand, the supply-side explanations of FLB suggest that the bias is more observed in bookmaker markets than in pari-mutuel markets, however, it cannot provide a clear explanation unless FLB present in pari-mutuel betting markets (Hurley and McDonough 2013; Metsola 2010).[7]

---

[6] Quandt (1986), assuming that bettors are risk lovers with mean variance utility functions, argues that long shots are expected to have greater subjective probabilities than objective probabilities, while favorites are expected to have smaller subjective probabilities. On the other hand, Golec and Tamarkin (1998) suggest that the long shot anomaly can be explained by bettor preference for return skewness rather than by preference for risk.

[7] There are two principal types of racetrack betting market: the pari-mutuel and bookmaker's system: Pari-mutuel means literally betting against other bettors, as opposed to betting against bookmakers. In the pari-mutuel system, first developed by Pierre Oller from France in 1865, the total amount of bets up until the start of a race is deducted by a fixed proportion of the pool to cover taxes and operating cost, and divided among the holders of the winning bettors. Bets with bookmaker are at a marginal price, whereas dividend paid by the pari-mutuel system are an average.

2.2 Exceptions to the FLB

Despites of the wide evidence of FLB, some researchers have found that the reverse favorite longshot bias (hereafter, RFLB) in Asian countries. The studies for Hong Kong (Busche and Hall, 1988; Busche, 1994;) and Japan (Walls and Busche, 2003) show Asian deviation from FLB.

The RFLB in Asian markets remains as a big puzzle of the field (Coleman 2004). One possible hypothesis is that the higher pool size and the higher expected return of Hong Kong racetrack market result in more careful and accurate bets (Benter 1994). According to Coleman (2004), the RFLB of Hong Kong is maybe resulted from its relatively large size of pool size or its tight regulation or due to distinctive characteristic of Chinese bettors. Some other researchers state that the Hong Kong race market bettors are more risk-adverse or risk-neutral bettors compared to Western race market bettors and the reason behind is cultural difference (Busche and Hall 1988). If RFLB is related to Asian culture, it is likely to show in Korean market either, as the cultural root of Korea is the same as Hong Kong and Japan. This justifies the motivation for an empirical investigation on FLB hypothesis in Korean race track market.

3. Overview of Korean Racetrack Market and Basic Concepts of racetrack betting

Horse races are run in three cities in Korea: Seoul, Jeju, and Busan. The five basic horse racing bets in Korea are Win bets, Place bets, Exacta bets, Quinella bets, and Quinella Place bets. The odds are determined differently depending on the types of bets. In this study, we will focus on Win bets which pays only when the horse finishes first in the race. Korea adopts pari-mutuel betting system.

Win pool of each race can be described as the sum of every bet on every horse running in the race. Racetrack odds refer to the actual return on a unit bet on the wining horse. The odds reflect how much a horse is favored to win. In the pari-mutuel system the odds are determined from the bets made by the public. Under the pari-mutuel system, the track operator extracts a percentage of the betting pool and returns the remainder to winning bettors in proportion to their individual stakes on the outcome of the race.

Suppose there are N horses and let B_i be the amount of money that has been bet for horse i to win. Then, Win pool (W) and the odds of each horse i ($Odds_i$) are given by the following equations (3.1) and (3.2) when track take is denoted by t.

$$W = \sum_{i=1}^{N} B_i \tag{3.1}$$

$$Odds_i = \frac{(1-t)W - B_i}{B_i} = (1-t)\frac{W}{B_i} - 1 \tag{3.2}$$

Payoffs of Win bets (the actual net return, NR_i) is determined by the odds. NR_i becomes the odds when betting on a horse to win, as only the bets for the winning horse pay off money. Otherwise, it is impossible to win any money; in this case, the return would be -1 as defined in equation (3.3)

$$NR_i = \frac{(1-t)W}{B_i} * 1_{[i=win]} - 1 \tag{3.3}$$

4. Model and Empirical Strategy

4.1 Risk Preference Model

The most common procedure to test for the existence of the FLB in risk preference hypothesis is comparing the subjective winning probabilities and the objective winning probabilities. Subjective probability contains the information that is the proportion of the total betting pool bet on each horse. If the market is efficient, the subjective probabilities should reflect the actual winning outcome, thus the subjective winning probability should be equal to each horse's objective win probability.

To test relationship between the objective winning probability (OP_{ij}) and the subjective probability (SP_{ij}) for a horse *i* participating in the *j*th race, the empirical model can be defined as equation (4.1)

$$SP_{ij} = a + bOP_{ij} + e_{ij} \tag{4.1}$$

In equation (4.1), the market efficiency hypothesis implies $H_0: a = 0,\ b = 1.$ In equation (4.1), the subjective probability of winning can be easily constructed after the race. If we ignore the track take for simplicity, following Griffith (1949), SP_{ij} can be written as $SP_{ij} = \frac{B_{ij}}{\sum_{i=1}^{N} B_{ij}}$. Then SP_{ij} can be written as (4.2) using (3.1) and (3.2).

$$SP_{ij} = \frac{1}{Odds_{ij}+1} \tag{4.2}$$

While SP_{ij} is calculated by win odds in equation (4.2), OP_{ij} is not observable. This problem has been an obstacle in testing the betting market efficiency. Since the objective probability of winning is not observable, equation (4.1) cannot be directly estimated.

In order to define the objective winning probability, many attempts have been made. Ali (1977) develops the method of grouping data to test the hypothesis of market efficiency. It is practically impossible to estimate the objective winning probability of a particular horse in a race, because the horse runs only once. A single observation of win or not would not be enough for a reasonable estimation of the winning probability. To overcome this problem, Ali (1977) assumes that the horses in the same 'odds rank' (or 'favorite') in different races are identical. In other words, Ali (1977) assumes that the winning probabilities of the horses in the same ‘favorite’ in different races are random picks from an ‘identical’ probability distribution. It would be equivalent to assuming a same set of horses run all the races repeatedly with a fixed ‘favorite’ rank. By doing so, Ali (1977) could estimate the objective winning probability of the horses by the actual winning ratio (= the number of wins / the number of races) of each ‘favorite’ group.

Ali’s method has been adopted by many researchers for various types of betting markets, including racetrack (Benter 1994; Terrell and Farmer 1996; Busche and Walls 2000; Gramm 2005; Gulati and Shetty 2007). Ali’s method, however, depends on an unrealistic and unverifiable assumption. In reality, even though two horses running different races have the lowest odds in each race, they could be heterogeneous in many ways. For example, they could be different in speed, age, or weight. It is not realistic to assume the winning probabilities of those two horses are from an identical probability distribution. It is more plausible to assume the heterogeneity of horses alter the probability of winning. Furthermore, the homogeneity assumption cannot be empirically verified, as each horse is observed only once in each race.

Our empirical strategy is more general than Ali’s grouping method. We assume that the objective probability of winning is, although it is not observable, a function of relevant covariates. The covariates would be horse-specific characteristics, such as its experience, age, weight, speed, or medical conditions. Of course, some race-specific characteristics related to the horse’s performance, such as race distance or time of the race could be included in the covariates. We further assume a linear relationship between the objective probability and the covariates as follows.

$$OP_{ij}^{*} = c + d'X_{ij} + u_{ij} \tag{4.3}$$

where OP_{ij}^* stands for the unobservable objective probability of horse i's winning in a race j, X_{ij} for the vector of covariates (both horse-specific and race-specific), and u_{ij} for a random disturbance. This framework is better than Ali (1977) in the sense that it allows the heterogeneity in the horses and the races.

Though we cannot fully observe the objective probability in practice, we do have a partial observability through a binary outcome coming from the objective probability. It is reasonable to assume that horse *i* wins in race *j* if its objective probability of winning is higher than a (unobservable) threshold. We define a binary outcome of 'win or not' as I_{ij} which takes a value of 1 if horse *i* won in race *j*, and a value of zero otherwise. The observability rule of the binary outcome is:

$$I_{ij} = \begin{cases} 1 & \text{if } OP_{ij}^* \geq \mu \\ 0 & \text{if } OP_{ij}^* < \mu \end{cases} \tag{4.4}$$

where μ is the threshold for winning.

If we combine equation (4.3) and observability rule (4.4), the following binary regression model is setup.

$$I_{ij} = c + d'X_{ij} + u_{ij} \tag{4.5}$$

Assuming a normal distribution for u_{ij}, we can consistently estimate the parameters *c* and d by so-called probit MLE (maximum likelihood estimation). Using the estimated parameters, we can also predict OP_{ij}^* in terms of observable covariates. Our strategy is to use the predicted value of OP_{ij}^* as the regressor of our main equation (4.1), and test for the market efficiency hypothesis, $H_0: a = 0 \; and \; b = 1$. Formally, we estimate the following model.

$$SP_{ij} = a + b\,\widehat{OP}_{ij}^{*} + e_{ij} \tag{4.6}$$

where $\widehat{OP}_{ij}^{*}$ is the predicted value of the objective winning probability for horse i in race j.

This method could be more generalized if we use the rank observations rather than binary 'win or not.' Suppose we have observed the ranks of top five horses in every race. Then the observability rule (4.4) will be generalized as follows.

$$I_{ij} = 5 \quad if\ OP_{ij}^{*} \geq \mu_1$$

$$I_{ij} = 4 \quad if\ \mu_2 \leq OP_{ij}^{*} < \mu_1$$

$$I_{ij} = 3 \quad if\ \mu_3 \leq OP_{ij}^{*} < \mu_2$$

$$I_{ij} = 2 \quad if\ \mu_4 \leq OP_{ij}^{*} < \mu_3$$

$$I_{ij} = 1 \quad if\ \mu_5 \leq OP_{ij}^{*} < \mu_4$$

$$I_{ij} = 0 \quad if\ OP_{ij}^{*} < \mu_5$$

Using this observability condition, we can estimate the multinomial regression model (4.5) by an ordered probit MLE, and compute the predicted objective probability of first-ranked in a similar way to the binary case.

4.2 Information Model

An alternative approach to test the FLB hypothesis is 'information model' comparing economic returns between different odds classes or favorite positions. Information model suggests a simple linear relationship between the objective and the subjective probability, whereas the risk preference model predicts a non-linear relationship between them. There are two types of information model. One type of study relies on Bayesian updating, while the other is based on fixed odds (Williams and Paton,1998; Shin, 1991; Terrell and Farmer, 1996).

To compare with the risk preference model suggested in section 4.1, we test the FLB hypothesis by information model, too. Following Williams and Paton (1998), we define our empirical model by equation (4.7).

$$NR_{ij} = \gamma_1 + \gamma_2 Odds_{ij} + \varepsilon_{ij} \tag{4.7}$$

If horse *i* wins the race j, its net return (NR_{ij}) equals its odds; otherwise, NR_{ij} is -1. The OLS (Ordinary Least Squares) estimation of the regression model is not appropriate because the dependent variable is left truncated at -1. Therefore, a Tobit MLE (Maximum Likelihood Estimation) is used to handle the truncation problem. If γ_2 is significantly lower than 0, then it means that the bets on small odds have high returns. Such result could be the evidence of FLB. However, if γ_2 is greater than 0, we can say FLB does not exist.

5. Empirical Results

The data employed in our study are the races run at Seoul Race Park, held between January 2012 and December 2013.[8] The data set is composed of 25,090 horses participated in 2,184 races during our data period. On average, 6 races are held in one racing day. The average race consists of 11 horses (minimum 6 to maximum 16). The total number of horses in our data

[8] We are grateful to KRA (Korea Racing Association) for providing the data.

period is 2,086 and each horse runs 12.03 races on average in the two years (minimum 1 to maximum 39). Among the 2,086 horses, 9 horses have run at least 30 times and 218 horses have run at least 20 times.

Table 1 gives the definitions and the descriptive statistics of all the variables in the empirical models. As is seen in Table 1, we use six variables for X_{ij} in equations (4.3) and (4.5): Speed, Popularity, Race Distance, Number of Horses, Age and Weight.[9]

Table 1. Description of Variables

Variable	Description	Sample Size	Mean	SD	Min	Max
Dependent Variables						
OP^*_{ij}	Latent variable representing the horse i's objective probability of winning the j^{th} race					
Win	If win =1, otherwise = 0	25090	0.09	0.28	0.00	1.00
Ranking	Ordered ranking categories from 0 to 5: 1st place (5), 2nd place (4), 3rd place (3), 4th place (2), 5th place (1) and no ranking (0)	25090	1.31	1.76	0.00	5.00
SP_{ij}	Subjective winning probability of horse i in race j	25090	0.085	0.096	0.0006	0.5
NR_{ij}	Net return of betting on horse i in race j	25090	-0.18	5.55	-1.00	280.10
Explanatory Variables						
X_{ij}						
Speed	Average speed (meter/second)	25090	15.29	0.42	11.28	16.69
Popularity	Popularity Rank of horses in each race (KRA provides a pre-expected popularity ranking of horse in each race) If 1, It is the most popular horse in each race.	25090	6.43	3.57	1.00	16.00
Race Distance	The total length over which the race will be run (meter)	25090	1406.21	292.74	1000.00	2300.00
Number of horses	The number of horses taking part in each race	25090	11.81	1.81	6.00	16.00
Age	The age of each horse which participates in the race	25090	3.56	1.02	2.00	10.00

[9] The variable 'Speed' is the average speed (= race distance divided by lap time record).

Weight	The weight of each horse which participates in the race	25090	53.72	1.76	48.00	64.00
$Pr(OP_{ij}^{*})$						
Pr(I=1)	Predicted value of objective winning probability derived by Probit model	25090	0.09	0.13	0.00	0.91
Pr(I=5)	Predicted value of objective winning probability derived by Ordered Probit model	25090	0.09	0.13	0.00	0.96
$Odds_{ij}$	The money of win bet on horse i in race j	25090	1 45.74	156.69	1.00	1566.90

Estimation results are presented in Table 2, Table 3 and Table 4. Table 2 is the result of estimating the objective probabilities with probit and ordered probit model. All explanatory variables are significant and the main results of explanatory variables for the ordered probit model (column 3) coincide with the basic binary probit model (column 1). We find that winning probability increases with the horse's speed and the race distance, while the probability decreases with the horse's popularity rank. The faster horse is more likely to win the race and the more popular horse has higher probability to win the race.[10] Age and weight of the horse have negative effects on the chance of winning, which indicates that the younger (lighter) the horse is, the more likely the winning probability is. It is also shown that the number of horses in a race tends to lower the winning probability of each horse.

Table 2. Result of Estimating Objective Winning Probability

	(1) Model 1(Probit)		(2) Model 2(Ordered Probit)	
	Coefficient	Std Error	Coefficient	Std Error
Speed	1.775***	0.053	2.085***	0.031
Popularity	-0.183***	0.006	-0.171***	0.003
Race Distance	0.0020***	0.000	0.002***	0.000
# of horses	-0.030***	0.007	-0.069***	0.004
Age	-0.070***	0.018	-0.076***	0.010
Weight	-0.024***	0.008	-0.010***	0.005
Constant	-28.79***	0.970		
Threshold1			32.500	0.562

[10] It is noted that the popularity variable is the expected rank of the horse. Thus, the lower the value is, the more popular the horse is.

Threshold2			32.836	0.563
Threshold3			33.201	0.563
Threshold4			33.631	0.564
Threshold5			34.220	0.565
Log-likelihood	-5348.62		-28173.62	
Pseudo R^2	0.279		0.1895	
# of observations	25090		25090	

Notes: Estimates in Model 1 and Model 2 are regression from probit and ordered-probit regression respectively.
The threshold points are the estimates of the threshold coefficients of the distribution function.
$*** P < 0.01$, $** P < 0.05$, $* P < 0.1$

Table 3 provides the marginal effects of the explanatory variables on the winning probability. They are calculated from the probit and ordered probit estimates, and evaluated at the sample means. The marginal effect of the horse's speed is quite large compared to other explanatory variables. For instance, one meter per second increase of the horse's speed raises the winning probability by 11 percent point according to the probit result, and by 12 percent point according to the ordered probit result. The other variables show relatively lower marginal effects, although all the marginal effects are statistically significant.

Table 3. Marginal Effects on Probability of Winning

	(1) Model 1(Probit)		(2) Model 2(Ordered Probit)	
	Coefficient	Mean	Coefficient	Std Error
Speed	0.1103***	0.0044	0.1168***	0.0034
Popularity	-0.0114***	0.0004	-0.0096***	0.0003
Race Distance	0.0001***	0.0000	0.0001***	0.0000
# of horses	-0.0019***	0.0005	-0.0039***	0.0003
Age	-0.0044***	0.0011	-0.0042***	0.0006
Weight	-0.0015***	0.0005	-0.0006***	0.0003
	y=Pr(win): 0.02692		y = Pr(i_rank=5): 0.02378	
# of observations	25090		25090	

Notes: Estimates in Model 1 and Model 2 are regression from probit and ordered-probit regression respectively.
The threshold points are the estimates of the threshold coefficients of the distribution function.
$*** P < 0.01$, $** P < 0.05$, $* P < 0.1$

Our principal focus is on Table 4. Table 4 layouts the relationship between the objective winning probability and the subjective probability in the race-track market of Korea. It is noted that the coefficients on the subjective probability are statistically significant in both model (1) and model (2). There is a positive relationship between objective and subjective probability. However, the coefficient is significantly lower than 1 in both the models. This result implies that the subjective probability does not fully reflect the true winning probability. Accordingly, we can conclude that the FLB exists in the racetrack betting market of Korea.

Table 4. Regression Result Estimating the relationship between Two Winning Probabilities

	(1) Model 1		(2) Model 2	
	Coefficient	Std Error	Coefficient	Std Error
Pr(I=1)	0.571***	0.00318		
Pr(I=5)			0.508***	0.003
Constant	0.035***	0.0005	0.041***	0.000
Adjusted R^2	0.5624		0.4966	
#of observations	25090		25090	

*** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$

This result is somewhat interesting. Unlike other Asian countries such as Hong Kong and Japan, Korean racetrack market shows the same characteristic as the western countries. With this result, we could speculate the reversed FLB is not a special Asian character. Rather, it can be explained as a country-specific phenomenon. Coleman (2004) tries to explain the absence of FLB in Hong Kong as an aspect of the country. Huge size of pools, small but well-established runners and Chinese gamblers propensity could be the reasons of the reversed FLB in Hong Kong.

As a comparison, Table 5 presents the estimation results of the information model using Tobit regression. In this case, the odds coefficient is found to be -0.397, which is highly significantly different from zero. The result coincides with the risk preference model indicating FLB in Korean racetrack market. Moreover, the significantly negative value of the constant meets the expectation that the net return would have a negative value after the taxes and commissions are deducted from the track take.

Table 5. Tobit Regression Result of Information Model

	Coefficient	Standard Error	T-value
Odds	-0.3970***	0.0187	-21.24
Constant	-30.168***	0.6855	-44.01

Sigma	27.8757	0.5052	-
Log Likelihood	-14276.93		
#of observations	25090		
Left censored at –1	22906		
Uncensored	2184		

$***P < 0.01$, $**P < 0.05$, $*P < 0.1$

6. Comparison to Traditional Method

The new methodology we propose above would be better than the traditional method of the estimated objective probability by Ali (1977) in three aspects. First, our method is not subject to the restrictive assumption that the winning probabilities of the horses in the same 'favorite' in different races are random picks from an 'identical' probability distribution. Second, the new method utilizes more information through the covariates of binary (or ordered categorical) regression. Third, as the traditional method dichotomizes the result as win or lose, our regression approach could be easily generalized to various qualitative categories such as win, place, exacta, quinella, etc.

In this section, we compare the actual prediction accuracy of our method to the traditional method. What we compare are three probabilities: 1) the objective probability of winning predicted by the traditional method, 2) the objective probability of winning predicted by our method, and 3) the actual probability of winning. We will see which prediction of 1) or 2) is closer to the actual probability 3).

For the comparison, we chose two subsamples of horses: the first group is the horses that have run more than 29 races in our data set, and the second group is the horses that have run more than 19 races in our data set. We have identified 9 horses in the first group, and 218 horses in the second group.[11] The reason why we chose the horses with as many races as possible is that the actual probability of winning (i.e. the number of wins / the number of races)

[11] Of course, the 9 horses in the first group are also included in the second group.

of a certain horse would become more precise as the horse runs more races. We compute the above three probabilities 1) – 3) for the 9 first group horses as follows.

1) First, we compute the expected winning probability of each 'favorite' of horses by Ali's method.[12] The results are presented at Table 6. Then we compute the 'frequency' of each horse being in each favorite, and multiply it to the expected winning probability of the favorite in Table 6. The sum of these 'frequency-weighted' expected probabilities of winning is the traditional (Ali's) prediction of objective winning probability. For example, let us assume that a particular horse, say Sunday Silence, ran 30 races in our data. In the 30 runs, Sunday Silence was rated as 1st favorite 5 times, 2nd favorite 15 times, and 3rd favorite 10 times. Then the predicted objective winning probability of Sunday Silence would be:

$$0.3827\times\frac{5}{30}+0.1998\times\frac{15}{30}+0.1207\times\frac{10}{30}=0.2039\,.$$

Table 6. Ali's Expected Winning Probability of Each Favorite

Favorite	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Prob	0.3827	0.1998	0.1207	0.0899	0.0666	0.0463	0.0336	0.0232	0.0141	0.0105	0.0063	0.0071	0.0033	0.0046	0.0000	0.0000

2) The computation of the predicted winning probability by our method is rather simple. From equation (4.3), the regression coefficients are estimated by binary probit or ordered probit. Using the estimated coefficients and the average observed covariates of each horse, we can compute the marginal effects and eventually the predicted probability of winning for each horse.

3) The actual probability of winning is computed by the total number of wins of the horse divided by the total number of races the horse ran in the observed period.

Table 7 below shows the above three probabilities 1) – 3) for the 9 horses that have run more than 29 races. It is clearly shown in Table 7 that our new method predicts the actual

[12] There exist 16 'favorites' in our data, as the maximum number of horses in a race was 16.

winning probabilities of these 9 horses more accurately than the traditional method. The last two columns present the squared prediction errors of the two methods. The sum of squared prediction errors (SSPE) by the traditional method is 0.007491, which is almost 6 times bigger than the SSPE by our new method, 0.001294.

Table 7. Prediction Accuracy Comparison of First Subsample

Horse	Number of Races	Number of Wins	Actual Prob. of Wining (A)	Traditional Prediction (B)	New Prediction (C)	(Trad.) Sq. Pred Error $(B-A)^2$	(New) Sq. Pred Error $(C-A)^2$
1	39	0	0	0.0309	0.0039	0.000952	0.000015
2	36	0	0	0.0254	0.0042	0.000644	0.000017
3	35	0	0	0.0412	0.0044	0.001699	0.000020
4	35	0	0	0.0202	0.0033	0.000406	0.000011
5	34	1	0.0294	0.0645	0.0124	0.001231	0.000288
6	31	0	0	0.0387	0.0072	0.001499	0.000052
7	30	1	0.0333	0.0345	0.0070	0.000001	0.000692
8	30	0	0	0.0123	0.0009	0.000152	0.000001
9	30	0	0	0.0301	0.0141	0.000907	0.000198
Sum of Squared Prediction Errors (SSPE)						0.007491	0.001294

As the first subsample of 9 horses could be too small for a fair comparison, we expand the prediction accuracy assessment to a bigger subsample: 218 horses that have run more than 19 races. We apply the same procedure explained above to the group of 218 horses and compare the SSPE's of the traditional method and our new method. The results are summarized in Table 8. As shown in Table 8, the SSPE of our new method, 0.625433, is much smaller than the SSPE of the traditional method, 1.069537. To summarize, the within-sample prediction accuracy comparison shows that our new method has better prediction power than the traditional method. As explained above, the fact that our method is not subject to the restrictive assumption of Ali (1977) and that the new method utilizes more information through the covariates of binary (or ordered categorical) regression would be the reason behind this accuracy gain.

Table 8. Prediction Accuracy Comparison of First and Second Subsample

	Number of Horses	SSPE of Traditional Method	SSPE of Our New Method
Subsample 1	9 horses	0.007491	0.001294
Subsample 2	218 horses	1.069537	0.625433

7. Conclusion

The FLB (favorite-longshot bias) is well recognized in the economic literature and a lot of empirical evidence have been presented from different countries. It can be described as 'one of the most robust anomalous empirical regularities in economics' (Walls and Busche 2003). However, there exist deviations from FLB (reversed FLB) in some Asian markets such as Hong Kong and Japan.

This paper investigates the anomalous phenomenon in racetrack betting market with respect to FLB and the market efficiency using Korean racetrack data. The traditional method comparing the subjective and the objective winning probability has an unavoidable limitation because it needs to proxy the unobservable true objective probability of winning. We propose a new method to forecast the objective probability using a latent variable approach. Our method is more general than any other previous approaches, in the sense that it can utilize many covariates reflecting the heterogeneity in the winning probability. In the within-sample prediction accuracy comparison, our method has shown a noticeably better performance.

Our empirical results from risk preference and information model support the existence of FLB in Korean race track betting market. By employing a new method that estimates the objective winning probability, we have shown that FLB exists in the race track market of Korea. We use binary and categorical regressions to determine the predicted values of objective probability and estimate the relationship between the predicted objective winning probability and the subjective probability. The results are consistent with the studies of Western countries, but it distinguishes Korean racetrack market from other Asian countries.

The existence of FLB in Korea can be a good start to solve the puzzle of the deviation of Asian racetrack market. Although it may be difficult to directly compare the race track market of Hong Kong or Japan with Korean market, the cultural similarity in the three countries would give us a path to navigate in the future.

References

Abinzano, I., Muga, L. and Santamaria, R., 2017. Behavioral biases never walk alone: an empirical analysis of the effect of overconfidence on probabilities. *Journal of Sports Economics, 18*(2), pp.99-125.

Ali, Mukhtar M., 1977, Probability, and utility estimates for racetrack bettors. *Journal of Political Economy, 85*(4), pp.803-815.

Benter, W., 1994. Computer-based horse race handicapping and wagering systems: A report. *Efficiency of racetrack betting markets*, pp.183-198.

Borghesi, R., 2012. The implications of a reverse favourite-longshot bias in a prediction market. *Journal of Prediction Markets, 6*(2), pp.12–21.

Brown, W.O., Sauer, R.D., 1993. Fundamentals or noise? Evidence from the professional basketball betting market. *The Journal of Finance, 48*(4), pp.1193–1209.

Bruce, A.C. and Johnson, J.E., 2000. Investigating the roots of the favourite-longshot bias: An analysis of decision making by supply-and demand-side agents in parallel betting markets. *Journal of Behavioral Decision Making, 13*(4), p.413.

Bruce, A.C. and Johnson, J.E., 1992. Toward an explanation of betting as a leisure pursuit. *Leisure Studies, 11*(3), pp.201-218.

Busche, K., 1994. Efficient market results in an Asian setting. In: Hausch, D.B., Lo, V.S., Ziemba, W.T. (Ed.), *Efficiency of Racetrack Betting Markets*. Academic Press, New York, pp.615–616.

Busche, K. and Hall, C.D., 1988. An exception to the risk preference anomaly. *Journal of business,* pp.337-346.

Busche, K. and Walls, W.D., 2000. Decision costs and betting market efficiency. *Rationality and Society, 12*(4), pp.477-492.

Cain, M., Law, D., Peel, D., 2000. The favourite-longshot bias and market efficiency in UK football betting. *Scottish Journal of Political Economy, 47*(1), pp.25–36.

Cain, M., Law, D., Peel, D., 2003. The favourite-longshot bias, bookmaker margins and insider trading in a variety of betting markets. *Bulletin of Economic Research, 55*(3), pp.263–273.

Camerer, C.F., 1989. Does the Basketball Market Believe in the Hot Hand,'?. *The American Economic Review, 79*(5), pp.1257-1261.

Coleman, L., 2004. New light on the longshot bias. *Applied Economics, 36*(4), pp.315–326.

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*(2), pp.295–318.

Deschamps, B. and Gergaud, O., 2012. Efficiency in betting markets: evidence from English football. *The Journal of Prediction Markets, 1*(1), pp.61-73.

Direr, A., 2013. Are betting markets efficient? Evidence from European football championships. *Applied Economics, 45*(3), pp.343–356.

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

Forrest, D., Simmons, R., 2000. Forecasting sport: the behaviour and performance of football tipsters. *International Journal of Forecasting, 16*(3), pp.317–331.

Forrest, D., Goddard, J., Simmons, R., 2005. Odds-setters as forecasters: the case of English football. *International Journal of Forecasting, 21*(3), pp.551–564.

Gandar, J.M., Zuber, R.A. and Johnson, R.S., 2001. Searching for the favourite-longshot bias down under: An examination of the New Zealand pari-mutuel betting market. *Applied Economics, 33*(13), pp.1621-1629.

Gandar, J.M., Zuber, R.A., Johnson, R.S., 2004. A reexamination of the efficiency of the betting market on National Hockey League games. *Journal of Sports Economics, 5*(2), pp.152–168.

Gandar, J.M., Zuber, R.A., O'Brien, T., Russo, B., 1988. Testing rationality in the point spread betting market. *Journal of Finance, 43*(4), pp.995–1008.

Golec, J. and Tamarkin, M., 1998. Bettors love skewness, not risk, at the horse track. *Journal of political economy, 106*(1), pp.205-225.

Golec, J., Tamarkin, M., 1991. The degree of inefficiency in the football betting market. *Journal of Financial Economics, 30*(2), pp.311–323.

Graham, I., Stott, H., 2008. Predicting bookmaker odds and efficiency for UK football. *Applied Economics, 40(*1), pp.99–109.

Gramm, M. and Owens, D.H., 2005. Determinants of betting market efficiency. *Applied Economics Letters, 12*(3), pp.181-185.

Gray, P.K., Gray, S.F., 1997. Testing market efficiency: evidence from the NFL sports betting market. *The Journal of Finance, 52*(4), pp.1725–1737

Gray, P.K., Gray, S.F., Roche, T., 2005. Efficiency of football betting markets: the economic significance of trading strategies. *Accounting & Finance, 45*(2), pp.269–281

Griffith, R.M., 1949. Odds adjustment by American horse race bettors, *American Journal of Psychology, 62*(2), pp.290-294.

Gulati, A. and Shetty, S., 2007. Empirical evidence of error in pricing of favorites and longshots in greyhound racing. *Journal of Economics and Finance, 31*(1), pp.49-58.

Hausch, D.B., Ziemba, W.T. and Rubinstein, M., 1981. Efficiency of the market for racetrack betting. *Management science, 27*(12), pp.1435-1452.

Hurley, W. and McDonough, L., 2013. Imperfect market-maker competition, heterogeneous expectations, and the favourite-longshot bias in wagering markets. *The Journal of Gambling Business and Economics, 1*(1), pp.3-12.

Hurley, W. and McDonough, L., 1995. A note on the Hayek hypothesis and the favorite-longshot bias in parimutuel betting. *The American Economic Review*, pp.949-955.

Jullien, B. and Salanié, B., 2000. Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy, 108*(3), pp.503-530.

Kim, J.Y., 2008. Testing for the favorite-longshot bias in racetrack betting: Evidence from Korean Horse Racing. Master's Thesis, Yonsei University

Koning, R.H., 2012. Regression tests and the efficiency of fixed odds betting markets. *International Journal of Sport Finance, 7*(3), p.262.

Lahvička, J., 2014. What causes the favourite-longshot bias? Further evidence from tennis. *Applied Economics Letters, 21*(2), pp.90-92.

Malkiel, B.G., 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives 17*, pp.59–82.

Metsola, M., 2010. Evidence on the Favorite-Longshot Bias as a Supply-Side Phenomenon. *Journal of Prediction Markets, 4*(1), pp.59-77.

Nyberg, H., 2014. A multinomial logit-based statistical test of association football betting market efficiency. *HECER Discussion Paper* No. 380. Helsinki Center of Economic Research, Helsinki.

Ottaviani, M. and Sørensen, P.N., 2010. Noise, information, and the favorite-longshot bias in parimutuel predictions. *American Economic Journal: Microeconomics, 2*(1), pp.58-85.

Paul, R. and Weinbach, A., 2005. Market efficiency and NCAA college basketball gambling. *Journal of Economics and Finance, 29*(3), pp.403-408.

Paul, R.J. and Weinbach, A.P., 2012. Sportsbook Pricing and the Behavioral Biases of Bettors in the NHL. *Journal of Economics and Finance, 36*(1), pp.123-135.

Pope, P.F. and Peel, D.A., 1989. Information, prices and efficiency in a fixed-odds betting market. *Economica*, pp.323-341.

Quandt, R.E., 1986. Betting and equilibrium. *The Quarterly Journal of Economics, 101*(1), pp.201-207.

Ro, Y.J., 2014. Searching for the 'Favourite-Longshot' Bias: Evidence from Korean Horse Racing Market. Master's Thesis, Yonsei University.

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), pp.173-190.

Shin, H.S., 1991. Optimal betting odds against insider traders. *The Economic Journal,* 101(408), pp.1179-1185.

Kanto, A.J., Rosenqvist, G. and Suvas, A., 1992. On utility function estimation of racetrack bettors. *Journal of Economic Psychology, 13*(3), pp.491-498.

Swidler, S. and Shaw, R., 1995. Racetrack wagering and the "uninformed" bettor: A study of market efficiency. *The Quarterly Review of Economics and Finance, 35*(3), pp.305-314.

Terrell, D. and Farmer, A., 1996. Optimal betting and efficiency in parimutuel betting markets with information costs. *The Economic Journal*, pp.846-868.

Thaler, R.H. and Ziemba, W.T., 1988. Anomalies: Parimutuel betting markets: Racetracks and lotteries. *The Journal of Economic Perspectives, 2*(2), pp.161-174.

Tuckwell, R.H., 1983. The thoroughbred gambling market: efficiency, equity and related issues. *Australian Economic Papers, 22*(40), pp.106-118.

Williams, L.V. and Paton, D., 1997. Why is there a favourite-longshot bias in British racetrack betting markets? *The Economic Journal*, pp.150-158.

Vlastakis, N., Dotsis, G. and Markellos, R.N., 2009. How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting, 28*(5), pp.426-444.

Walls, W. D. and Busche, K., 2003. Broken odds and the favourite-longshot bias in parimutuel betting: a direct test. *Applied Economics Letters, 10*(5), pp.311-314.

Weitzman, M., 1965. Utility analysis and group behavior: An empirical study. *Journal of political Economy, 73*(1), pp.18-26.

Williams, L.V. and Paton, D., 1997. Why is there a favourite-longshot bias in British racetrack betting markets. *The Economic Journal,* pp.150-158.

Williams, L.V. and Paton, D., 1998. Why are some favourite-longshot biases positive and others negative?. *Applied Economics, 30*(11), pp.1505-1510.

Williams, L.V., 1999. Information efficiency in betting markets: A survey. *Bulletin of Economic Research, 51*(1), pp.1-39.

Winter, S. and Kukuk, M., 2006. Risk Love and the Favorite-Longshot Bias: Evidence from German Harness Horse Racing. *Schmalenbach Business Review (sbr), 58*(4), pp.349-364.

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), pp.269-279.

Woodland, L.M. and Woodland, B.M., 2001. Market efficiency and profitable wagering in the national hockey league: Can bettors score on longshots? *Southern Economic Journal*, pp.983-995.