2008

Market Efficiency at the Derby: A Real Horse Race

Steven D. Dolvin
Butler University, sdolvin@butler.edu

Mark K. Pyles

Market Efficiency at the Derby: A Real Horse Race

Steven D. Dolvin
Assistant Professor of Finance, School of Business and Economics
Butler University, E-mail: sdolvin@butler.edu

Mark K. Pyles
Assistant Professor of Finance, School of Business and Economics
College of Charleston, E-mail: pylesm@cofc.edu

Abstract:

Using race data from each Kentucky Derby from 1920 to 2005, we examine whether the horse wagering market is efficient. Most prior studies in this arena test potential betting strategies that rely on posted odds, generally finding that it is extremely difficult to devise and implement any consistently successful wager (i.e., market efficiency). We extend these studies by examining underlying determinants of posted race odds, specifically focusing on the experience of auxiliary members (e.g., jockey, breeder and trainer) associated with each entrant. We find that past Derby experience is an important determinant of posted odds and that the odds-making system appears to capture relevant experience, as using this data provides no incremental information for forming market-beating wagers. Thus, we provide additional evidence supporting efficiency in horse race betting markets.

I. Introduction

By definition, market efficiency implies that relevant information is compounded quickly into posted prices, and, therefore, market participants cannot earn a consistently positive excess return simply by forming strategies based on this known set of information. The majority of market efficiency tests focus explicitly on markets for financial securities such as stocks and bonds, generally concluding that financial markets are, at least, not inefficient. However, any market that exhibits similar characteristics should also be subject to the concept of market efficiency.[1]

For example, financial markets can be characterized by the following conditions: uncertain returns, numerous profit-seeking participants, and an extensive set of available information. Obviously, these characteristics are not unique to financial markets. In fact, Ali (1998) contends that these are all relevant descriptions of the horse wagering market. Further, Quandt (1986) suggests that the horse wagering market exhibits other similarities to the financial markets. Specifically, the profitability of investors depends on both objective (skill of firm managers or horse jockeys) and subjective (what other investors or bettors think) factors.

[1] Sauer (1998) presents a comprehensive survey on the economics of wagering markets, including horse racing.

Given the similarities between horse racing and securities markets, multiple studies have examined the efficiency of wagering on horse races by attempting to identify systematic deviations that allow for the creation of consistently successful betting strategies.[2] For example, Rosett (1965, 1971), Snyder (1978), and Ali (1979) find that participants tend to overbet long-shots and underbet favorites. Further, McGlothlin (1956) and Asch, Malkiel, and Quandt (1982) find the tendency to overbet long-shots is strongest in late races.

Asch, Malkiel, and Quandt (1984) extend the above studies by examining morning line odds, finding that profits cannot be consistently earned in win betting, but it may be possible to exploit these inconsistencies in show or place betting.[3] However, Asch, Malkiel, and Quandt (1986) reverse this latter claim and conclude it is unlikely that any potential strategy could consistently generate excess profits. They conclude that bettors, as a whole, are rational and the market is efficient. This is consistent with the previous findings of Snyder (1978). Thus, the general consensus is that the horse wagering market, similar to financial markets, is efficient.[4]

The previous literature identified above typically takes posted odds as given; however, it is possible that some information generally thought to be controlled for in the odds making system may not be fully incorporated. If this is the case, then it may be possible to earn an excess return even if existing studies that focus explicitly on final posted odds (or, equivalently, the associated probability of winning) suggest otherwise. Thus, we extend existing studies by evaluating the impact of some determinants of posted odds in an effort to see if the odds fully capture certain pre-race knowledge. Specifically, we examine the effect of prior Derby experience of three of the main (non-horse) players in a race: (1) the jockey, (2) the breeder, and (3) the trainer.

We employ a unique sample with which to examine market efficiency in the horse wagering market: the Kentucky Derby. We examine all Derby entrants from 1920-2005, specifically focusing on experience of the supporting members for each horse. As suggested above, our unique data, combined with our new approach, has several potential contributions to the literature. First, we are unaware of previous work that examines specific determinants of posted race odds. Second, most prior studies have examined data sets at lesser known tracks over multiple races. We examine the most famous horse race in the United States, avoiding many potential biases that occur due to time of race differentials, different track lengths and sizes, and extent of publicity. In this, we create a more consistent data set over time, which reduces, for example, potential biases associated with early- versus late-race betting. Third, following Asch, Malkiel, and Quandt (1984) and Ali (1998), we extend the results of existing studies that examine horse racing as a test of market efficiency. Specifically, we take a closer look at the

[2] Plackett (1975) and Henery (1981) are among the first to examine the probability of winning in a horse race. Most papers use these results as a catalyst for examining market efficiency.

[3] "Win" betting refers to a wager that attempts to identify the winner of the race. "Place" ["Show"] betting refers to selecting a horse that will either win or place (2nd) [win, place (2nd) or show (3rd)]. A horse that finishes in the top three spots is also referred to as finishing "in the money."

[4] Even studies that find systematic deviations suggest that implementing the strategy is difficult, if not impossible. Thus, even if the market is not fully efficient, it is, at a minimum, transactionally efficient.

primary variable of interest in previous studies (i.e., odds) by examining how and to what extent experience influences odds.

We use a variety of methods to measure the experience of each player, subsequently applying a two-stage approach to control for potential endogeneity embedded in the data. We find that experience is indeed an important determinant in creating post odds. In fact, it appears that experience may be, other than the horse itself, the most important determinant. Further, it does appear that posted odds fully capture experience, as no significant relations remain between experience and winning (or finishing in the money) after we control for it via our two-stage approach. Our results are consistent with previous studies that find efficiency in the market, implying that posted odds (similar to financial market prices) appear to be a robust estimation of the horse's potential to win the race.

II. Data

We examine each Kentucky Derby entrant for every race from 1915 to 2005. We obtain our data from the Kentucky Derby media guide created by Churchill Downs. The data are also available online at *www.kentuckyderby.com*. Within the media guide are the Derby charts, which contain place of finish for each entrant, along with a variety of information about the race (e.g., times, post positions, and track conditions).[5]

We also compile data for each entrant's jockey, breeder, and trainer.[6] As our primary focus is to examine the impact of previous Derby experience of these auxiliary participants on the race outcome, we create several variables to measure experience. As our primary sample begins with the 1920 Derby, we use the years 1915 to 1919 as a reference point for measuring experience. Therefore, the participant is coded as having previous experience if they had participated (in the capacity judged) in any prior Derby, beginning in 1915. We examine four measures of experience for each participant. The first, *PriorDum*, is a binary variable equal to one if the horse entrant was ridden (jockey), bred (breeder), or trained (trainer) by an individual who had at least one previous entrant in the Kentucky Derby, zero otherwise.

In order to examine the extent of experience, we also create *PriorNumb*, which is defined as the number of prior entrants ridden, bred, or trained by the jockey, breeder, or trainer, respectively. If the participant had multiple horses in the same race (in the case of the breeder or trainer), both were counted as previous experience; therefore, *PriorNumb* is not necessarily the number of prior Derbies in which the breeder or trainer

---

[5] The authors would like to thank Ms. Cathy C. Schenck of the Keeneland Assocation for assistance locating and compiling the data used in this study.

[6] We also examine the horse owner; however, there is considerable overlap between the breeder and the owner, particularly until recent years. Therefore, we choose to exclude this participant from our primary analysis. In unreported results, however, we examine the entire study in reference to the owner and find results qualitatively equivalent to those for the breeder.

participated.[7] Further, we examine the *success* of prior experience with *PriorWin* and *PriorMoney*. *PriorWin* is a binary variable equal to one if the horse's jockey, breeder, or trainer had previously won a Derby, zero otherwise. *PriorMoney* is a binary variable equal to one if the respective participant had previously been associated with an entrant that finished in the top three, zero otherwise.

As controls, we also extract horse-type variables. Specifically, we control for geldings and fillies. Geldings are entrants that, strictly speaking, have been castrated, while fillies are female horses that have yet to reach sexual maturity. Both may have an effect on the probability of winning, as the vast majority of Derby entrants are neither fillies nor geldings.[8]

The horse's post position may also affect its potential outcome, as much attention is paid to the draw in the week prior to the Derby. For example, there have been more winners from post positions 1(12), 4(10), and 5(12) than any other position. However, this is likely influenced by the number of horses in the race (i.e., posts 1-5 would always have entrants, whereas post 20 would not). Therefore, instead of using the number of the post position as a control, we split the post position variable into three segments; (1) *inside*, (2) *mid*, and (3) *outside*. If the number of positions is divisible by 3, each segment receives an equal number of horses. For example, if there are 9 entrants, post positions 1, 2, and 3 will be characterized as having *inside* post positions, while 4, 5 and 6 will be *mid*, and 7, 8, and 9 will be *outside*.

If the number is not equally divisible, we adjust as follows. If there are n-1 entrants, where n is an equally divisible number, we split the sample into (n/3, (n/3)-1, n/3). If there are n-2 entrants, we split the sample as ((n/3)-1, n/3, (n/3)-1). In addition, we control for the field size, defined as the number of entrants in each Derby, as well as the condition of the track.[9]

Rather than using posted odds, we transform our primary independent variable into the probability of winning, which is a more consistent variable across races, particularly when different field sizes exist. We follow Ali (1998) by defining a particular entrant's probability of winning as follows:

$$probability_i = \frac{\frac{1}{1+O_i}}{\sum_{i=1}^{n}\frac{1}{1+O_i}} \qquad (1)$$

[7] To examine potential nonlinear relations, in unreported results we examine *PriorNumb* squared, but we find no significance relative to the results reported.

[8] Specifically, of the 1,305 entrants, only 90 were geldings (i.e., 6.9%), while only 21 were fillies (i.e., 1.7%). Of those, only 3 geldings and 2 fillies won the derby over the period examined.

[9] The condition of the track is constant with respect to each entrant in a particular race. However, it is widely understood that certain horses tend to run better in certain conditions (e.g., sloppy or muddy tracks). This should be factored into the odds in an efficient market. Therefore, we control for this in our regressions. In unreported results, we eliminate the track condition controls from the regressions and find the results unchanged.

where O_i is each entrants posted odds. As such, for each Derby the sum of the probabilities is 1.

III. Results

We begin by examining summary statistics that measure the level and impact of the experience of each participant of interest: jockey, breeder, and trainer. We evaluate the level of experience using the metrics defined earlier (i.e., *PriorDum*, *PriorNumb*, *PriorWin*, and *PriorMoney*). Further, we give specific attention to whether experience is associated with a higher occurrence of winning (i.e., placing 1^{st}) or finishing in the money (i.e., placing 1^{st}, 2^{nd}, or 3^{rd}). Results are presented in Table 1.

Panel A reports the relation between experience and winning, where *WinDum* is defined as a binary variable equal to 1 if the entrant wins the Derby in which it was entered, zero otherwise. Panel B examines the relation between experience and finishing in the money, where *MoneyDum* is a binary variable equal to one if the entrant finished in the money in the Derby in which it was entered, zero otherwise.

Examining the results, it appears that horses ridden by jockeys with prior experience, regardless of definition, are more likely to win, as well as place in the money. We find similar results with regard to the breeder and trainer. The only exception is that breeders that had previously bred a horse that finished in the money are not necessarily more likely to win. However, taken as a whole, it appears that previous Derby experience certainly matters, at least at a univariate level. The question thus becomes whether or not this influence is fully captured in posted odds (i.e., does the market efficiently reflect information related to experience levels?).

To address this question, we extend the analysis by examining the influence of relevant variables, including experience, on posted odds. However, as defined above, rather than evaluating odds explicitly, we convert posted odds into the underlying probability of winning. Papke and Wooldridge (1996) suggest when analyzing a dependent variable whose values are constrained between zero and one, which is the case with posted probabilities, the most appropriate statistical approach is a fractional logit model. We follow this approach and present the results of the following model in Table 2:[10]

[10] For robustness, we also examine our results using a standard OLS approach. Technically speaking, there is little difference between using a logit model and traditional OLS. The main advantage of the fractional model, which is relevant to our analysis, is the predicted values are constrained between 0 and 1, while OLS structurally does not dictate this. However, in our sample, all predicted values from the OLS approach are between zero and one. Thus, even after revising the statistical approach, our results are qualitatively the same, which adds robustness to our findings.

$$\begin{aligned} Probability = \beta_0 + \beta_1 JockeyEx + \beta_2 BreederEx + \beta_3 TrainerEx + \beta_4 Gelding + \beta_5 Filly + \\ \beta_6 Inside + \beta_7 Outside + \beta_8 FieldSz + \beta_9 GoodDum + \beta_{10} HeavyDum + \\ \beta_{11} MuddyDum + \beta_{12} SlowDum + \beta_{13} SloppyDum + \varepsilon \end{aligned} \tag{2}$$

Probability is the calculated probability of winning as defined in eq. (1). *JockeyEx*, *BreederEx*, and *TrainerEx* are experience variables as represented by each participant's respective *PriorDum* (Column 1), *PriorNumb* (Column 2), *PriorWin* (Column 3), or *PriorMoney* (Column 4).[11] *Gelding*, *Filly*, *Inside*, and *Outside* are as defined previously. *FieldSz* is the total number of entrants in the respective Derby. *GoodDum, HeavyDum, MuddyDum, SlowDum,* and *SloppyDum* are binary variables equal to one if the track is in each respective condition at post time, zero otherwise.[12]

Examining the results in Table 2, we find a negative relation between the size of the field and the horse's probability of winning. This is expected in that a larger field makes, presumably, for a more competitive race, and, therefore, each horse has a lower relative probability of winning. In addition, geldings are negatively associated with the probability of winning, which is consistent with the low number of geldings entered into Derbies over the sample period. The low level of participation is likely related to the perceived ineffectiveness of horses with this specific characteristic, which would be manifest in a lower probability of winning. None of the other secondary variables of interest are highly significant, although inside post position does have a moderately significant (10 percent level througout) positive relation to the probability of winning.[13]

We next turn to our primary variables of interest, i.e., the experience measures. We find each experience measure (i.e., *PriorDum*, *PriorNumb*, *PriorWin*, and *PriorMoney*) for jockeys, breeders, and trainers has a consistently significant and positive relation to the entrant's probability of winning the Derby. This indicates that odds are contingent, at least somewhat, on the experience of the auxiliary members associated with each horse. In fact, given the significance in relation to the other control variables, it appears that prior experience may be the most important determinant (other than the

[11] An obvious concern is potential correlation between the ancillary members' measures of experience used in each model. For example, if there is a high degree of correlation, then multicollinearity could result in inefficient estimates as the significance levels would be inflated. However, an examination of the correlation matrix of each set of variables indicates low levels of correlation (never exceeding .26). Therefore, it is unlikely that multicollinearity has much of an effect. Nonetheless, for robustness we redefine the models including each of the experience variables independently and find our primary results are qualitatively unchanged.

[12] The excluded variables for post position and track condition are *Middle* and *Fastdum*, respectively.

[13] It is possible that our results may be contingent upon the time period studied. For example, in the late 1980s, large horse races, such as the Derby, began to be broadcast to the public for wagering purposes (i.e., simulcast). Therefore, more people had the opportunity to place wagers on the outcome of the race. In order to examine if this is an important determinant in our study, we create variables to control for the time periods, one from 1985 to 1994 (the period where simulcasting began to gain in popularity) and the other from 1995 to 2005 (the period where simulcasting became widespread). However, including these two variables does not change the primary results. The same is true if we include a time trend variable for the entire time period. We thank an anonymous reviewer for this suggestion.

horse itself) of odds, and in-turn, the probability of winning. This remains true for all external members (i.e., the jockey, breeder, and trainer).[14]

Given that we have established a significant relation between the experience of the jockey, breeder, and trainer and the posted odds, we now examine the deeper question of market efficiency. Most previous work has attempted to do this by examining the odds in relation to the results, but they have not examined individual determinants of these odds to see if they are completely captured by the posted odds. We therefore wish to extend previous analyses by examining the second stage equation (i.e., logit) as follows:

$$Result = \alpha + \beta_1 PredProb + \beta_2 JockeyEx + \beta_3 BreederEx + \beta_4 TrainerEx + \varepsilon \qquad (3)$$

where *Result* is either *WinDum* or *MoneyDum*.[15] *PredProb* is the predicted probability of winning as calculated using the results in Table 2 (i.e., the first stage regression) for each experience measure, which allows us to control for potential endogeneity between posted odds and our experience measures. The experience variables are as defined above.

If the experience (at least as we define it) of the jockey, breeder, and trainer is fully captured in the posted odds (i.e., the predicted probability), then we expect to see no significance for the experience variables in this second stage regression. If significance remains, it is indicative of market inefficiencies, as the market has not fully processed all publicly available information and reflected such in the price (i.e., the odds) of the asset (i.e., the horse). Results are presented in Table 3.[16]

Examining Table 3, we find little-to-no significance in any of the experience measures in relation to *WinDum*. This finding indicates that, on average, experience of the auxiliary members of the horse team has been fully captured by the posted odds, and there is no consistent strategy that can be employed to "beat the market" by examining this information. Therefore, consistent with previous studies, it appears the horse wagering market is efficient.

---

[14] Obviously, the quality of the horse would be the key factor. And, it is likely that endogeneity exists in that the best horses are able to attract the most experienced jockeys. However, we have no available method for judging the quality (or experience) of particular horses, as horses only race in a single Derby. Further, we explored using a horse's prior race record; however, without a way to standardize race results across time, tracks, and competition, the use of such records are extremely limited. Thus, the experience, in addition to reflecting increased ability, may also proxy for information related to the horse itself. Further, we have no information on the horse's bloodline, which could also serve as a proxy for quality.

[15] The finishing position of each entrant is available. However, money is typically only earned on the first three horses. Exceptions are bets such as superfectas, which require the bettor to choose the first four finishing horses in order. However, our analyses do not focus on combination bets such as these, but rather on single horse bets. Therefore, we only examine horses that win or place in the money. In unreported results, we examine horses that place (i.e., finish 2nd) or show (i.e., finish 3rd) individually and find our result are unchanged in that we find no significance in the experience measures.

[16] The Logit model, which we use for examining eq. (3), possesses the independence of irrelevant alternatives property, which fits horse racing since the relative odds depend only on the characteristics of the particular horses. Further, Bacon-Shone, Lo, and Busche (1992) find that a logit model best fits this type of data.

The same is true when examining the top three finishers in each Derby, with one notable exception. We find a remaining positive and significant relation between a jockey with a prior Kentucky Derby ride and his mounted horse finishing in the money. This perhaps indicates that the jockey (who has the most in-race control of the three auxiliary members examined) can use his experience to guide a horse through the field slightly better than those with no experience. In other words, perhaps he can maximize the finish of a non-winning caliber horse, whereas an inexperienced jockey cannot.

As a simple test of the potential impact of this finding, we consider a particular scenario. Specifically, the findings suggest, for example, that a horse who is picked *ex ante* to place fourth should have a greater chance of finishing in the money if the jockey has prior Derby experience. Thus, for each race, we rank the entrants in descending order based on the calculated probability of winning. For each entrant with the fourth highest probability of winning, we identify whether the jockey has a prior Derby mount. We then test whether those with experience are more likely to finish in the money. However, the difference is insignificant, which suggests that even though there is a small statistical significance, the economic implication is small. Thus, overall, it appears that the market is, at least, transactionally efficient in relation to money horses as well as winners.

Although our primary concern has been addressed by examining market efficiency in relation to experience, for robustness we also examine the other variables used as controls in Table 2. If markets are efficient (which is the working hypothesis) then none of the other variables should have significant relations to *Result* in the second stage. Therefore, we examine the following expanded second stage model:

$$\begin{aligned} Result = {} & \alpha + \beta_1 PredProbPriorDum + \beta_2 GoodDum + \beta_3 HeavyDum + \beta_4 MuddyDum + \\ & \beta_5 SlowDum + \beta_6 SloppyDum + \beta_7 Gelding + \beta_8 Filly + \beta_9 Inside + \beta_{10} Outside + \\ & \beta_{11} JockeyEx + \beta_{12} BreederEx + \beta_{13} TrainerEx + \varepsilon \end{aligned} \tag{4}$$

For parsimony, we choose to report only results from the *PriorDum* analysis. The results are presented in Table 4. Naturally, in unreported results we examine the other three experience measures as well, but the results are qualitatively identical to those reported.

As there are three additional categories of explanatory variables here, we first examine each of them separately. In column 1 we examine track condition variables, while in columns 2 and 3 we examine horse type and post position variables, respectively. In column 4, we examine all variables combined, along with the experience variables for each auxiliary player.

The results support our previous findings of market efficiency. There is a unanimous lack of significance in all explanatory variables except the predicted probability of winning as calculated from column 1 of Table 2. Again, the only exception is the prior experience of the jockey in regards to the horse finishing in the top 3. Thus, our results as a whole appear to be robust, which further strengthens the findings of previous work that concludes the horse race wagering market is efficient.

IV. Conclusion

We examine horse racing odds for the Kentucky Derby in an effort to determine whether betting markets are efficient with regard to available pre-race information. We extend previous studies by examining the determinants of posted odds, rather than taking them as given. Specifically, we examine the impact of track conditions, horse type, post position, and auxiliary players' experience on the probability of winning the Derby. We find these experience measures may be the most predictive in creating post odds for each entrant. Using multiple experience measures for the jockey, breeder, and trainer, we find a positive relationship between prior experience and the probability of winning.

We then examine market efficiency by implementing a two-stage approach, finding no significant impact on the race outcome of experience (or any other explanatory variable) remains after controlling for its impact in determining posted odds. We interpret this as further evidence, following Snyder (1978) and Asch, Malkiel, and Quandt (1986), of market efficiency in that no consistent excess return can be generated based upon publicly available information.

These results have interesting, but disappointing, implications for bettors. On a broad scale, our results are consistent with the semi-strong form of market efficiency. In other words, any information that is publicly available appears to have already been incorporated into market prices (or odds in this case), and, therefore, no excess return can be generated, no matter the effort exerted by the investor (bettor) to extract and identify such information. This does not necessarily indicate strong-form efficiency, as we have no way to define and examine the influence of private (or inside) information, which could be described in the horse racing world as a "hot tip." This would be the next logical step of examination, should one find a way to isolate and identify "private" information.

References

Ali, Mukhtar M. (1979). "Some Evidence of the Efficiency of a Speculative Market," *Econometrica* 47: 387 - 392.

Ali, Mukhtar M. (1998). "Probability Models on Horse-Race Outcomes," *Journal of Applied Statistics*, 25 (2): 221 - 229.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt (1982). "Racetrack Betting and Informed Behavior," *Journal of Financial Economics*, 10: 187 - 194.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt (1984). "Market Efficiency in Racetrack Betting," *Journal of Business*, 58 (2): 165 - 175.

Asch, Peter, Burton G. Malkiel, and Richard E. Quandt (1986). "Market Efficiency in Racetrack Betting: Further Evidence and a Correction," *Journal of Business*, 59 (1): 157 - 160.

Bacon-Shone, J., V. Lo, and K. Busche (1992). "Logistic Analyses for Complicated Bets," *Research Report 11*, Department of Statistics, University of Hong Kong.

Henery, R. J. (1981). "Permutation Probabilities as Models for Horse Races," *Journal of the Royal Statistical Society: Series B (Methodological)*, 43 (1): 86 - 91.

McGlothlin, William H. (1956). "Stability of Choices Among Uncertain Alternatives," *American Journal of Psychology*, 69: 604 - 615.

Papke, Leslie E. and Jeffrey M. Wooldridge (1996). "Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates," *Journal of Applied Econometrics*, 11(6): 619 - 632.

Plackett, R. L. (1975). "The Analysis of Permutations," *Applied Statistician*, 24: 193 - 202.

Quandt, Richard E. (1986). "Betting and Equilibrium," *Quarterly Journal of Economics*, 101 (1): 201 - 208.

Rosset, Richard N. (1965). "Gambling and Rationality," *Journal of Political Economics*, 73 (December): 595 - 607.

Rosset, Richard N. (1971). "Weak Experimental Verification of the Expected Utility Hypothesis," *Review of Economic Studies*, 38: 481 - 492.

Sauer, R. D. (1998). "The Economics of Wagering Markets," *Journal of Economic Literature*, 36: 2021 – 2064.

Snyder, Wayne W. (1978). "Horse Racing: Testing the Efficient Markets Model," *Journal of Finance*, 33 (4): 1109 - 1118.

Table 1: Summary Statistics

The following table presents descriptive statistics for Kentucky Derby entrants from 1920 to 2005. Panel A examines *WinDum*, which is a binary variable equal to one if the entrant placed first in the respective Kentucky Derby, zero otherwise. Panel B examines *MoneyDum*, which is a binary variable equal to one if the entrant placed first, second, or third in the respective Kentucky Derby, zero otherwise. The rows in each panel examine the percentage of jockeys, breeders, and trainers, respectively, that win (i.e., *WinDum* in Panel A) or finish in the money (i.e., *MoneyDum* in Panel B). The columns are sorted by various measures of prior Derby experience, starting with the 1915 Derby. Specifically, *PriorDum* examines whether the jockey, breeder, or trainer, respectively, had any Derby experience prior to the sample entrant. *PriorNumb* examines the extent of experience, where we examine experience in at least two previous derbies versus those that had either zero or one previous Derby. *PriorWin* examines whether the jockey, breeder, or trainer in question had won a Derby prior to the sample entrant. *PriorMoney* examines whether the jockey, breeder, or trainer had placed first, second, or third in any previous Derby. For example, the first row/first column in Panel A indicates that 8 percent of jockeys with prior Derby experience win their races, whereas only 4 percent without experience win. The difference, which is tested in the third column of each section, is statistically significant, suggesting prior experience is a significant determinate of Derby performance. The remaining entries are interpreted similarly. *t*-statistics are calculated assuming unequal variances. Data are from www.kentuckyderby.com.

Panel A:

	PriorDum			*PriorNumb*			*PriorWin*			*PriorMoney*		
	Yes	No	*t*-stat	>=2	1 or 0	*t*-stat	Yes	No	*t*-stat	Yes	No	*t*-stat
JockeyEx	.08	.04	3.31	.09	.04	3.94	.10	.06	2.48	.10	.05	3.11
n	(855)	(450)		(642)	(663)		(278)	(1,027)		(447)	(858)	
BreederEx	.09	.05	2.39	.09	.06	1.73	.14	.06	2.67	.08	.06	1.21
n	(516)	(789)		(342)	(963)		(137)	(1,168)		(276)	(1,029)	
TrainerEx	.09	.05	2.59	.11	.05	3.43	.13	.05	3.41	.11	.05	3.21
n	(611)	(694)		(392)	(913)		(194)	(1,111)		(351)	(954)	

Panel B:

	PriorDum			*PriorNumb*			*PriorWin*			*PriorMoney*		
	Yes	No	*t*-stat	>=2	1 or 0	*t*-stat	Yes	No	*t*-stat	Yes	No	*t*-
JockeyEx	.24	.12	5.43	.26	.13	6.04	.27	.18	3.13	.28	.16	4.8
n	(855)	(450)		(642)	(633)		(278)	(1,027)		(447)	(858)	
BreederEx	.25	.16	3.60	.25	.18	2.46	.33	.18	3.49	.24	.19	1.8
n	(516)	(789)		(342)	(963)		(137)	(1,168)		(276)	(1,029)	
TrainerEx	.25	.15	4.46	.27	.17	3.92	.30	.18	3.41	.27	.17	3.7
n	(611)	(694)		(392)	(913)		(194)	(1,111)		(351)	(954)	

Table 2: Stage 1

The following table presents fractional logit results from the equation:

$$Probability = \alpha + \beta_1 JockeyEx + \beta_2 BreederEx + \beta_3 TrainerEx + \beta_4 Gelding + \beta_5 Filly + \beta_6 Inside$$
$$+ \beta_7 Outside + \beta_8 FieldSz + \beta_9 GoodDum + \beta_{10} HeavyDum + \beta_{11} MuddyDum + \beta_{12}\, SlowDum +$$
$$+ \beta_{13} SloppyDum + \varepsilon_i$$

where *Probability* is the entrant's calculated probability of winning based upon posted odds. *JockeyEx, BreederEx*, and *TrainerEx* are the primary variables of interest and correspond to the experience measure used in each regression. Specifically, Column 1 uses *PriorDum* to measure experience, while Column 2 uses *PriorNumb* and Columns 3 and 4 use *PriorWin* and *PriorMoney*, respectively. *Gelding* is a binary variable equal to one if the entrant was a gelding, zero otherwise. *Filly* is a binary variable equal to one if the entrant was a filly, zero otherwise. *Inside* is a binary variable equal to one if the entrant's post position is one of the inside third of the starting grid, zero otherwise. *Outside* is a binary variable equal to one if the entrant's post position is one of the outside third of the starting grid, zero otherwise. The excluded category is *Mid*. *FieldSz* is the number of horses in each Derby field. *GoodDum, HeavyDum, MuddyDum, SlowDum,* and *SloppyDum* are binary variables equal to one if the track at post is judged to be good, heavy, muddy, slow, or sloppy, respectively. The excluded category is *FastDum*. Data are from www.kentuckyderby.com.

	(1) *PriorDum*		(2) *PriorNumb*		(3) *PriorWin*		(4) *PriorMoney*	
	Coef.	*p*-val	Coef.	*p*-val	Coef.	*p*-val	Coef.	*p*-val
Intercept	-1.95	.01	-1.67	.01	-1.71	.01	-1.83	.01
JockeyEx	.29	.00	.05	.01	.46	.00	.44	.00
BreederEx	.14	.03	.02	.03	.20	.07	.19	.02
TrainerEx	.35	.00	.01	.12	.33	.00	.39	.00
Gelding	-.26	.01	-.24	.01	-.24	.01	-.24	.01
Filly	.08	.66	.18	.35	.10	.63	.09	.60
Inside	.13	.10	.15	.06	.14	.07	.15	.05
Outside	.09	.24	.08	.29	.08	.34	.08	.32
FieldSz	-.08	.03	-.08	.02	-.07	.03	-.08	.02
GoodDum	.04	.70	.01	.90	.01	.94	.09	.41
HeavyDum	.23	.19	.28	.09	.21	.18	.23	.14
MuddyDum	.05	.71	.02	.89	.05	.75	.07	.61
SlowDum	-.00	.99	.06	.66	.01	.97	.06	.68
SloppyDum	.02	.86	.01	.94	.01	.91	.01	.96
N	1,305		1,305		1,305		1,305	
Pseudo. R-Sq	.1256		.1377		.1459		.1650	

Table 3: Stage 2 (Experience Measures)

The following table presents logit regression results from the equation:

$$Result=\alpha+\beta_1 PredProb+\beta_2 JockeyEx+\beta_3 BreederEx+\beta_4 TrainerEx+\varepsilon_i$$

where *Result* is *WinDum* (Panel A) or *MoneyDum* (Panel B). *WinDum* is a binary variable equal to one if the entrant placed first in the Derby, zero otherwise. *MoneyDum* is a binary variable equal to one if the entrant placed first, second, or third in the Derby, zero otherwise. *PredProb* is the predicted probability of winning as calculated using the results in Table 2 for each experience measure. *JockeyEx, BreederEx*, and *TrainerEx* are the primary variables of interest and correspond to the experience measure used in each regression. Data are from www.kentuckyderby.com.

Panel A: WinDum

	(1) *PriorDum*		(2) *PriorNumb*		(3) *PriorWin*		(4) *PriorMoney*	
	Coef.	*p*-val	Coef.	*p*-val	Coef.	*p*-val	Coef.	*p*-val
Intercept	-4.21	.00	-3.74	.00	-3.79	.00	-3.82	.00
PredProb	15.02	.00	14.98	.00	13.55	.01	13.48	.01
JockeyEx	.47	.11	-.01	.71	.09	.76	.26	.35
BreederEx	.27	.27	-.01	.81	.41	.21	-.09	.74
TrainerEx	.02	.93	.03	.07	.35	.25	.31	.27
N	1,305		1,305		1,305		1,305	
% Concordant	64.2		62.1		64.1		65.8	

Panel B: MoneyDum

	(1) *PriorDum*		(2) *PriorNumb*		(3) *PriorWin*		(4) *PriorMoney*	
	Coef.	*p*-val	Coef.	*p*-val	Coef.	*p*-val	Coef.	*p*-val
Intercept	-3.06	.00	-2.55	.00	-2.65	.00	-2.69	.00
PredProb	17.87	.00	16.89	.00	17.94	.00	18.00	.00
JockeyEx	.44	.01	-.01	.63	-.11	.58	.16	.37
BreederEx	.20	.18	-.01	.57	.29	.20	-.08	.67
TrainerEx	.02	.89	.02	.25	-.01	.95	-.05	.80
N	1,305		1,305		1,305		1,305	
% Concordant	66.0		61.8		61.1		64.1	

Table 4: Stage 2 (All Measures)

The following table presents results from the equation:

$$Result = \alpha + \beta_1 PredProbPriorDum + \beta_2 GoodDum + \beta_3 HeavyDum + \beta_4 MuddyDum + \beta_5 SlowDum + \beta_6 SloppyDum + \beta_7 Gelding + \beta_8 Filly + \beta_9 Inside + \beta_{10} Outside + \beta_{11} JockeyEx + \beta_{12} BreederEx + \beta_{13} TrainerEx + \varepsilon_i$$

where *Result* is *WinDum* (Panel A) or *MoneyDum* (Panel B). *WinDum* is a binary variable equal to one if the entrant placed first in the Derby, zero otherwise. *MoneyDum* is a binary variable equal to one if the entrant placed first, second, or third in the Derby, zero otherwise. *PredProbPriorDum* is the predicted probability of winning as calculated with the results from column 1 in each panel of Table 2. *GoodDum, HeavyDum, MuddyDum, SlowDum,* and *SloppyDum* are binary variables equal to one if the track is in each respective condition at post time, zero otherwise. The excluded category is *FastDum*. *Gelding* and *Filly* are binary variables equal to one if the entrant was a gelding or filly, respectively, zero otherwise. *Inside* is a binary variable equal to one if the entrant's post position is one of the inside third of the starting grid, zero otherwise. *Outside* is a binary variable equal to one if the entrant's post position is one of the outside third of the starting grid, zero otherwise. The excluded category is *Mid*. *JockeyEx, BreederEx*, and *TrainerEx* are the primary variables of interest and correspond to prior experience as measured by *PriorDum*. Data are from www.kentuckderby.com.

Panel A: WinDum

	(1)		(2)		(3)		(4)	
	Coef.	*p-val*	Coef.	*p-val*	Coef.	*p-val*	Coef.	*p-val*
Intercept	-4.06	.00	-3.99	.00	-3.98	.00	-4.13	.00
PredProbPriorDum	19.50	.00	18.90	.00	19.17	.00	14.15	.01
GoodDum	.06	.87					.09	.82
HeavyDum	.10	.92					.06	.96
MuddyDum	.06	.87					.06	.92
SlowDum	.05	.92					.09	.87
SloppyDum	-.05	.92					-.04	.94
Gelding			-.34	.58			-.41	.51
Filly			.15	.84			.12	.88
Inside					.08	.77	.11	.68
Outside					-.24	.41	-.21	.48
JockeyEx							.48	.11
BreederEx							.29	.24
TrainerEx							.03	.90
N	1,305		1,305		1,305		1,305	
% Concordant	63.4		63.6		63.7		65.1	

Panel B: MoneyDum

	(1)		(2)		(3)		(4)	
	Coef.	*p-val*	Coef.	*p-val*	Coef.	*p-val*	Coef.	*p-val*
Intercept	-2.95	.00	-2.95	.00	-2.86	.00	-3.03	.00
PredProbPriorDum	22.07	.00	22.30	.00	22.16	.00	18.38	.00
GoodDum	.05	.85					.06	.81
HeavyDum	.08	.90					.07	.91
MuddyDum	.04	.90					.05	.88
SlowDum	.04	.92					.09	.79
SloppyDum	-.02	.96					-.02	.95
Gelding			.06	.86			.01	.98
Filly			-.71	.27			-.70	.28
Inside					-.06	.73	-.03	.88
Outside					-.21	.25	-.18	.32
JockeyEx							.43	.02
BreederEx							.20	.19
TrainerEx							.02	.90
N	1,305		1,305		1,305		1,305	
% Concordant	64.9		65.0		64.8		66.3	