

**An Economic Analysis of Horseracing's Pick Three Wager:
The Public's Misperception of the Binomial Distribution**

Gregory M. Ellis

Department of Economics
University of Washington

I. Introduction

In the age of the internet and intertrack satellite simulcasting, horseracing is enjoying a remarkable surge in popularity. Simulcast wagering is making larger wagering pools available to thousands of horseplayers across the United States. Fans from New York to California can legally bet on races taking place at virtually any track in the country. With increased accessibility to national betting pools has also come the increased availability of exotic betting opportunities that include the exacta, trifecta, superfecta, pick six, and the subject of this paper, the pick three. Where such exotic bets had once been novelties to complement standard win, place, and show wagering, they now dominate betting menus at most tracks.

The object of the pick three wager is simple: correctly choose the winner of three consecutive races. At many tracks, including Southern California's premier venue, Santa Anita, horseplayers may bet the pick three on any three consecutive races. For example, on a race day with nine races, seven pick three wagers will be available (Races 1-3, 2-4, ..., 7-9). My interest in the pick three wager is both personal and professional. As a casual horseplayer, I am interested in wagers amenable to economic and statistical analysis that might give me an edge over the general betting public not engaging in such analysis. In the foreword to Steven Davidowitz's excellent book *Betting Thoroughbreds*, renowned handicapper Andrew Beyer notes that many bettors approach racing as a quest for the Rosetta Stone, for a system that will explain the mysteries of horseracing. To some degree, I suppose I am guilty of that quest, but my interest in analyzing the pick three wager is also professional. As an economist, I am fascinated by the risk-averse betting strategies I see employed by the very risk-loving segment of the population that one observes at the racetrack. Moreover, my general approach to horseracing is rooted in the economic belief in efficient markets. With all of the handicapping information available to the public, from the old standard, the *Daily Racing Form*, to the plethora of handicapping sites on the internet, my view is that, for the most part, the win odds in each race accurately reflect the relative likelihood of horses winning their respective races. The goal of this research is to discover the extent to which most of the general betting public misplays the pick three wager, not

from the perspective of incorrectly handicapping a particular race, but from the perspective of misunderstanding the relevant underlying economics and statistics of the wager. The goal, in short, is to discover betting strategies that are not used as much as they should be; the goal is to identify wagers that pay more than they should in expected value terms, even assuming that the public is correctly handicapping each individual horserace.

The focus of this analysis will be on the number of pre-race betting favorites that are included on typical pick three wagering tickets¹. Historically, betting favorites win their races approximately one-third of the time. This statistic has been shown to be robust to racing venue, decade, and type of race. The data on which the empirical analysis of this paper is based, the 230 races at the Fall 2000 meet at Santa Anita, are no exception. Favorites won their respective races exactly 30 percent of the time. In the next section of the paper, the profile of winning pick three wagers is explored with the aid of the binomial distribution. In particular, the likelihood of winning pick three tickets with zero, one, two, or three betting favorites is examined theoretically and empirically. In section III, the theoretical connection between the pick three wager and a three-race win-parlay bet is established (accounting for differences in pari-mutuel takeout rates). The pick three wager is shown, actuarially, to be a more promising one than its win-parlay analogue. In section IV, simple linear regressions (and nonlinear ones that correct for serial correlation) demonstrate the superiority of the pick three wager over the win-parlay counterpart, and they identify the way in which the public systematically overplays favorites when placing pick three wagers. Section V briefly discusses the role of handicapping when implementing a successful pick three strategy before presenting concluding comments.

¹ When one makes a pick three wager, one does not know with certainty, of course, which horses will be the favorites in the second and third legs of the three-race sequence. However, the established morning-line odds found in the racing program are a very strong predictor of the eventual pre-race favorites in those legs.

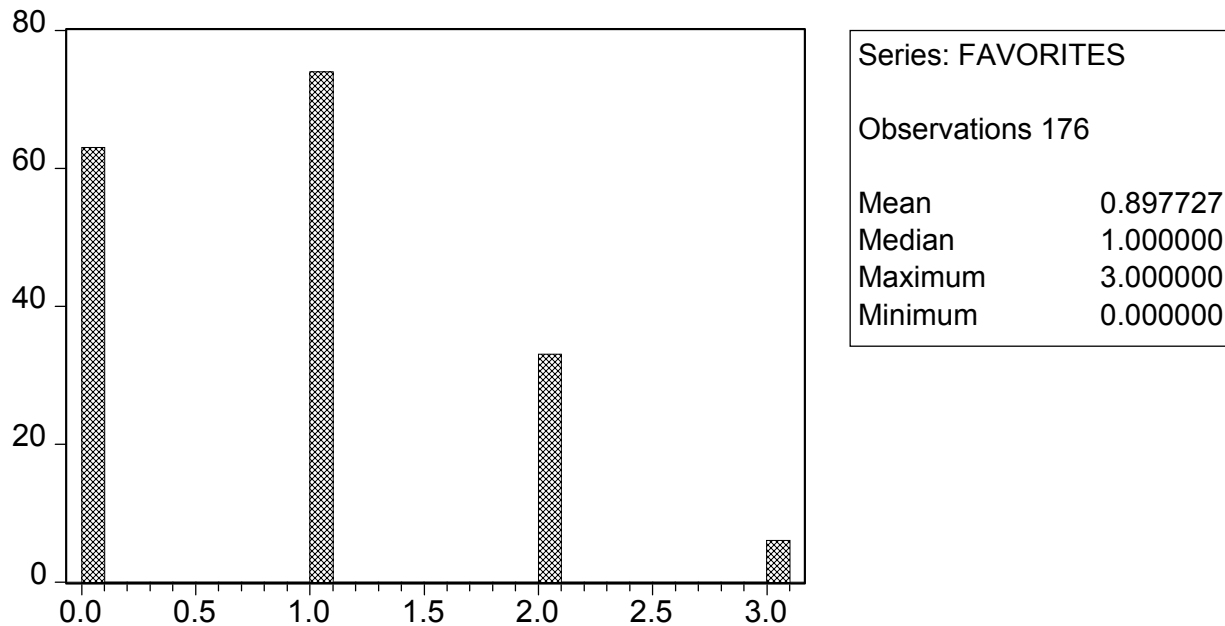
II. The Pick Three Wager and the Binomial Distribution

Pre-race betting favorites either win their races or they do not. Favorites at Santa Anita won 30 percent of their races during the Fall of 2000. Letting X be the number of favorites on a winning pick three ticket, the probability distribution of X is given by the binomial. In general, where X is the number of successes in n independent trials, the probability of observing x

successes is $f(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$ for $x=0, 1, \dots, n$. With $p=0.30$, $q=1-p=0.70$, and $n=3$, the theoretical probability distribution for the number of betting favorites on a winning pick three ticket is given in the table below:

x	0	1	2	3
$f(x)$	0.34	0.44	0.19	0.03

These theoretical probabilities are borne out in the data for the 176 pick three wagering opportunities at the Fall 2000 meet at Santa Anita. The histogram for the number of betting favorites on winning pick three tickets is presented below:



The number of winning tickets with zero betting favorites was 63, or 36%. The number of winning tickets with exactly one betting favorite was 74, or 42%. The number of winning tickets with exactly two favorites was 33, or 19%. Finally, the number of winning tickets with favorites winning all three races was 6, or 3%. Despite the low probability of favorites winning at least two of the three races in question (22%), we can demonstrate (in section IV of the paper) that the public places a large number of bets, pinning their hopes of success, on exactly that kind of unlikely profile with regard to betting favorites. In short, it is very rare for all three favorites to win their races in a particular pick three sequence, and when they do, the pick three payout is quite small (indicative of the larger number of folks making such wagers). In order to give further meaning to notions of small and large pick three payouts, we establish an actuarial benchmark in the next section of the paper: the win-parlay payout.

III. The Pick Three and Win-Parlay Payouts

Operating racetracks is an expensive venture. Between providing the stalls and workout facilities for the horses, the purses for the owners, trainers, and jockeys, and maintaining the track facility itself, a large sum of revenue must be collected from the horseplaying patrons to cover costs; it is. Each betting pool is subject to a particular pari-mutuel takeout rate. At Santa Anita, approximately 15 percent is taken from the conventional win, place, and show pools. Approximately 20 percent is taken off of the top of the exotic betting pools such as the exacta, trifecta, and pick three. This means that in a given race, only 85 percent of the total amount bet in the win pool is available to be paid out to those gambling on the winning horse. The sobering fact for participants in a pick three betting pool is that only 80 percent of the total amount bet is available to compensate those gamblers who correctly identify the winners of three consecutive races. Still, the pick three is widely considered to be one of the best wagers available at the track because, although one is essentially betting on three races, the takeout only occurs once.

To see the merit of this argument, consider the following example of a three-race win-parlay bet, the logical alternative to the pick three. Although the minimum win bet is typically \$2 at most tracks, imagine the possibility of a \$1 win bet on a horse in Race 1 with even-money odds. If the horse wins, one would win \$1 in addition to receiving back the original \$1 investment. One could then take the \$2 accumulated total and reinvest (parlay) it in Race 2. Suppose again that a winning bet is made on an even-money horse. Then, at the end of Race 2, one would have \$4 available for reinvestment in Race 3. If once again, a successful even-money win bet were made in Race 3, the \$1 would have grown to an \$8 win-parlay total.

In the example, the odds on each winning horse were even-money (one-to-one). With even-money win odds on a horse, what is the public's collective assessment of the probability that such a horse will actually win? The answer is less than 50%. Suppose that, in each race, \$100,000 is wagered in the win pool. After the 15% track takeout, the remaining pool available for payouts to the betting patrons is only \$85,000. To generate the even-money odds posted at the track, \$42,500 must have been wagered on the horse in question, implying that \$57,500 must have been wagered on all of the other horses in the field. If there was not a track takeout, the "true" win odds on the horse in question would be greater: approximately 1.35-to-one ($57,500/42,500$). In the previous paragraph, a betting example was described that yielded an \$8

win-parlay total after three races by assuming that the posted win odds of each winning horse were even-money. The same \$8 win-parlay total can be generated, mathematically, by employing the true win odds of 1.35-to-one rather than the even-money posted track odds. Specifically, by wagering 85% of the amount bet at each stage in the example, albeit at the more favorable true odds of 1.35-to-one, the same winning total of \$8 is generated at the end of the three race sequence. Thinking in terms of true win odds, rather than posted track win odds, will be useful in predicting expected pick three wager payouts (relative to win-parlay payouts).

Assuming the public's collective assessment of the likelihood of our three horses winning their respective races is accurate, what should the equivalent \$1 pick three wager on these horses pay, given that the takeout rate is higher, but that the takeout occurs only once? Because the takeout percentage is approximately twenty, it is as though for every dollar bet on a pick three wager, only 80 cents is actually wagered at the true win odds of 1.35-to-one in Race 1. That is, after Race 1, the successful pick three bettor in our example has theoretically accumulated \$1.88 to be reinvested in Race 2. After a successful reinvestment at true odds of 1.35-to-one, it is as though \$4.42 has been accumulated after Race 2. Finally, after successfully reinvesting the entire \$4.42 in Race 3, the bettor should expect to receive back a total of \$10.40. The ratio of the pick three payout to the three-race win-parlay payout is 1.30. With takeout rates of 15% in the win pools and 20% in the pick three pools, the premium to betting the pick three rather than the win-parlay should be 30%. This 30% theoretical benefit to playing the pick three is independent of the assumed even-money win odds (1.35-to-one true odds) in the example; the 30% premium does, of course, depend critically on the empirical facts that, in the example, win bet takeout rates are 15% and that the pick three takeout rate is 20%. To be more precise, at Santa Anita, the win bet takeout rate is 15.43% and the pick three takeout rate is 20.19%. Consequently, one can easily show that the premium to participating in the pick three wager at Santa Anita, rather than a win-parlay one, ought to be exactly 32%. In the Appendix, found at the end of the paper, we derive a formula to calculate the theoretical premium to making a pick three wager for any specified takeout rates on pick three or win bets.

The benchmark as to how successfully the public is employing various strategies in the pick three wager is this 32% premium. With data on the win odds for each winning horse in the 230 races at Santa Anita's Fall 2000 meet, we can calculate the resulting three-race win-parlay payouts, and compare them to the corresponding pick three payouts. If, for certain betting

strategies, the public is receiving less, on average, than this 32% premium over the corresponding win-parlay bet, then the public is overusing those strategies (resulting in disappointing payoffs relative to the expected premium). The pick three wager, like any other at the track, is a pari-mutuel one. The bettor is competing, not against the “house” as in Las Vegas table games, but against the other bettors participating in the same wagering pool. The track managers are guaranteed their 15 or 20 percent of the betting handle, regardless of the outcome of the races. Given the zero-sum nature of the competition between betting participants over the remaining money in the betting pools, a successful wagering strategy over time is necessarily one that is not used as often by the general public as the relevant true win odds would dictate. In the next section, we examine the success of various pick three wagering strategies with regard to the number of betting favorites included in the wager.

IV. An Empirical Analysis of Pick Three Wagering Strategies

Before formally analyzing wagering strategies through regression analysis, let us consider a few sample statistics for the 176 pick three wagers at Santa Anita. The average payout on a successful \$1 wager was \$317. The median payout was \$134, with a maximum payout of \$6,613 and a minimum payout of \$3. For the 63 winning pick three wagers involving zero pre-race favorites, the average and median payouts were \$597 and \$278 respectively. For the 74 winning wagers with exactly one favorite, the average and median payouts were \$212 and \$125. For the 33 winning wagers involving exactly two favorites, the average and median payouts were \$72 and \$50. Finally, the average and median payouts when all three favorites won their races were only \$16 and \$14 respectively, and this circumstance only prevailed 6 times. Other potentially relevant statistics from the Santa Anita meet include the average and median win odds for the winning horses in the 230 races analyzed: 4.8 and 3.2 to one respectively. Also, the average number of horses in a race, the average field size, was slightly more than 8.

Defining $ODDS1$, $ODDS2$, and $ODDS3$ as the win odds of the victorious horse in each of the three legs in a pick three sequence, the corresponding three-race win-parlay payout on a \$1 wager is

$$PARLAY3 = \{(ODDS1+1)*ODDS2 + (ODDS1+1)\} * ODDS3 + \{(ODDS1+1)*ODDS2 + (ODDS1+1)\}$$

Defining $PAYOUT$ as the pick three payout for the corresponding three-race sequence, we can regress $PAYOUT$ against $PARLAY3$ in order to verify the premium to betting the pick three versus the win-parlay:

$$PAYOUT_t = \beta * PARLAY3_t + u_t \quad (1)$$

The regression results are presented in Table 1. For every \$1 increase in win-parlay payout, we find a corresponding \$1.37 increase in pick three payout. This result is approximately within a standard error of our theoretical prediction that there ought to be a 32% premium to the pick three payout over the win-parlay payout.

However, this simple linear regression is suspect. The regression errors may very well be subject to problems of second-order serial correlation. For instance, if a particular pick three wager pays much better than the corresponding win-parlay bet would predict (due, for example, to a winning horse in the third leg being played much less in the pick three pool than the win odds would suggest), that positive regression error may persist over the course of the next two pick three wagers (during the same race day) because those wagers involve overlapping races. In other words, we may face an error structure of the form

$$u_t = \rho_1 * u_{t-1} + \rho_2 * u_{t-2} + \varepsilon_t \quad (2)$$

To test for the presence of second-order serial correlation, we employ the Breusch-Godfrey Lagrange Multiplier method, so the auxiliary regression involves regressing the current pick three regression residuals against *PARLAY3* and the previous two observed residuals. The dataset is intentionally constructed in such a way that distinct days of racing are separated by blank rows; consequently, only 122 of the original 176 pick three observations can be used in the auxiliary regression. Under the null hypothesis (of no serial correlation), the LM test statistic for the presence of second-order autocorrelation has a Chi-Square distribution. The test statistic is 10.33, and the critical value at the 5 percent level of significance is 5.99. The null hypothesis is rejected.

Table 2 contains the regression results for the quasi-differenced model accounting for second-order autocorrelation:

$$\begin{aligned} &PAYOUT_t - \rho_1 * PAYOUT_{t-1} - \rho_2 * PAYOUT_{t-2} \\ &= \beta * (PARLAY3_t - \rho_1 * PARLAY3_{t-1} - \rho_2 * PARLAY3_{t-2}) + \varepsilon_t \end{aligned} \quad (3)$$

The estimation is by nonlinear least squares, and it is performed with the EViews econometric software. The reported coefficients for the AR(1) and AR(2) terms are the estimated first and second-order autocorrelation coefficients, ρ_1 and ρ_2 respectively. Note that the regression model appears to be subject to positive serial correlation as we previously surmised. The

estimated pick three premium (relative to the win-parlay wager) is 34%, and the hypothesis that it is the theorized 32% cannot be rejected.

That the win odds for a horse are a sufficient summary statistic for likely exotic bet payoffs is sometimes questioned by horseplayers who insist that other variables like the size of the field in a race might help explain payoffs. Of course, the counterargument is that, to the extent that field size matters, it is already reflected in the relevant win odds. A regression that allows us to test this hypothesis in connection with the pick three wager is presented in Table 3. Defining *FIELDSIZE* as the total number of horses racing in the three-race sequence, our Table 2 regression is augmented by its inclusion as an explanatory variable for pick three payoffs. The t-statistic associated with the coefficient on *FIELDSIZE* is a mere 0.86, so the hypothesis that *FIELDSIZE* has no explanatory power cannot be rejected. At least as far as the pick three wager is concerned, win odds (and the corresponding win-parlay payouts) seem to be a sufficient statistic for likely payouts.

Table 4 presents ordinary least squares regressions of *PAYOUT* against *PARLAY3* for subsamples of the data regarding the number of favorites, *NUMBERFAV*, on the winning pick three ticket². Note that the highest premium to be found relative to the win-parlay bet, 80%, is when exactly one favorite wins its race in the three-race sequence. Apparently, the betting public is systematically ignoring this profile when wagering on the pick three, and pick three wagering strategies involving exactly one betting favorite are discussed in more detail in the next section of the paper.

In examining Table 4, it is clear that the lowest premium relative to the win-parlay bet, 16%, occurs when all three favorites win their races in the three race sequence. The premium is also subpar when favorites win exactly two of the three races in a pick three sequence. This is consistent with behavior in which the bettor is willing to take a chance on a longshot in one race, but too risk-averse to leave out the favorites in the other two races. A subset of this case occurs when a horse other than the favorite wins the first leg, and favorites win the second and third legs of the sequence. This pick three outcome only occurred 11 times at the Santa Anita meet, but the regression presented in Table 5 suggests a very small premium, 14%, in this event. My

² Corrections for serial correlation are not practical in these subsample regressions given the limited number of observations for each regression.

casual observation consistent with this low premium is that horseplayers analyzing the upcoming race will often identify a horse they like at decent odds, but because they have not yet comprehensively analyzed the two races subsequent to the upcoming one, they will invariably include the morning line favorites for those two races on their pick three tickets, even if the favorites are vulnerable to defeat. If enough horseplayers behave in this somewhat risk-averse fashion, the payouts will be disappointing. Including favorites on pick three tickets can be viewed as purchasing insurance, albeit incomplete insurance; after all, the favorite is *more* likely than any other horse to win a given race. But that does not mean that the favorite is *likely* to win the race. Actuarially speaking, given the 20% takeout rates on pick three wagering pools, the racetrack is a terrible place to purchase insurance!

V. The Role of Handicapping

The lesson learned from the Table 4 regressions should be that one can conceivably prosper from pursuing pick three wagering strategies that use at most one favorite in any three-race sequence. In fact, it appears as though the most promising strategies involve using exactly one favorite on a given pick three ticket. This begs the question, which favorite? Moreover, to suggest that a bettor participating in the pick three would do well to throw out at least two favorites is all very fine, but then the question becomes, which of the remaining seven or eight other horses in a race should he or she consider using? The answer to the latter question is beyond the scope of this paper (and, arguably, the reasoning abilities of the author), but some previously published research can shed some light on the former question as to which favorites to throw out or include in a pick three wager.

Our underlying assumption throughout the preceding analysis has been that, in establishing the win odds for horses in a particular race, horseplayers collectively establish the correct relative prices. That is, with all of the handicapping information available today to horseplayers competing in a national simulcast setting, it is hard to “beat the market”. For the most part, I continue to believe this to be the case. However, handicapper Steve Unite, a regular columnist for *Horseplayer Magazine*, recently examined a random sample of 326 races to specifically address the question of factors that help favorites win or lose more often than the norm. Factors that help a favorite win more often than the norm include, the race in question being the third race for the horse off of a layoff, the horse dropping from the category of maiden special weight to maiden claimer, the horse rising up the maiden claiming ranks, the horse attempting a route (more than eight furlongs) for the first time, or the horse receiving the medication Lasix for the first time. On the other hand, Unite found that factors making a favorite particularly vulnerable to defeat include the horse dropping to its lowest career claiming price or, most important of all, the horse being a favorite despite possessing a last-race speed figure that was not among the top 3 in the field. With this kind of evidence in mind, a horseplayer contemplating a pick three wager might try to find a three-race sequence in which two of the likely favorites are vulnerable for the aforementioned reasons, but a sequence in which the remaining favorite looks particularly formidable.

Pick three strategies involving exactly one betting favorite are relatively affordable in that, with a betting favorite singled in a particular leg of the pick three, the bettor can reasonably include several non-favorite horses on the ticket in the other two legs, without increasing the cost of the ticket too much. For example, a \$1 pick three wager that uses two horses in Race 1, one horse in Race 2, and three horses in Race 3 only costs the bettor \$6. The premium in payouts that these one-favorite strategies offer has been found to be substantial in our regression analysis.

The horseracing game is about risk and reward. This analysis has found that, for the risks taken, horseplayers too often engage in pick three wagering strategies that rely heavily on betting favorites. Our analysis shows that horseplayers intent on pursuing multiple-favorite strategies could well be heard to cry, “we don’t win very often, but at least the payout is lousy.”

Table 1

Dependent Variable: PAYOUT
Method: Least Squares
Included observations: 176

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.374830	0.047822	28.74876	0.0000
R-squared	0.784761	Mean dependent var	316.9949	
Adjusted R-squared	0.784761	S.D. dependent var	660.3293	

Table 2

Dependent Variable: PAYOUT
Method: Least Squares
Included observations: 122
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.344704	0.050229	26.77166	0.0000
AR(1)	0.218871	0.062959	3.476389	0.0007
AR(2)	0.084177	0.048132	1.748854	0.0829
R-squared	0.872321	Mean dependent var	293.8943	
Adjusted R-squared	0.870175	S.D. dependent var	488.4816	

Table 3

Dependent Variable: PAYOUT
Method: Least Squares
Included observations: 122
Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.328962	0.053402	24.88596	0.0000
FIELD SIZE	0.758532	0.878973	0.862975	0.3899
AR(1)	0.208594	0.063208	3.300108	0.0013
AR(2)	0.079441	0.048219	1.647511	0.1021
R-squared	0.873099	Mean dependent var	293.8943	
Adjusted R-squared	0.869873	S.D. dependent var	488.4816	

Table 4

Dependent Variable: PAYOUT
Method: Least Squares
Sample: IF NUMBERFAV=0
Included observations: 63 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.340214	0.079592	16.83849	0.0000
R-squared	0.753755	Mean dependent var	596.7619	
Adjusted R-squared	0.753755	S.D. dependent var	985.8178	

Dependent Variable: PAYOUT
Method: Least Squares
Sample: IF NUMBERFAV=1
Included observations: 74 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.804366	0.064845	27.82579	0.0000
R-squared	0.876197	Mean dependent var	212.4351	
Adjusted R-squared	0.876197	S.D. dependent var	323.5807	

Table 4 (continued)

Dependent Variable: PAYOUT
Method: Least Squares
Sample: IF NUMBERFAV=2
Included observations: 33 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.283729	0.066283	19.36751	0.0000
R-squared	0.828543	Mean dependent var	72.05758	
Adjusted R-squared	0.828543	S.D. dependent var	67.32707	

Dependent Variable: PAYOUT
Method: Least Squares
Sample: IF NUMBERFAV=3
Included observations: 6 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.159014	0.110097	10.52720	0.0001
R-squared	0.882669	Mean dependent var	16.16667	
Adjusted R-squared	0.882669	S.D. dependent var	13.51172	

Table 5

Dependent Variable: PAYOUT

Method: Least Squares

Sample: IF FAV1=0 AND FAV2=1 AND FAV3=1

Included observations: 11 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PARLAY3	1.137976	0.062726	18.14208	0.0000
R-squared	0.942492	Mean dependent var	87.90000	
Adjusted R-squared	0.942492	S.D. dependent var	94.57066	

References

Breusch, T., "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, vol. 17, 1978, pp. 334-355.

Davidowitz, Steven, *Betting Thoroughbreds: A Professional's Guide for the Horseplayer*, A Plume Book, published by the Penguin Group, 1997.

Godfrey, L. G., "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables," *Econometrica*, vol. 46, 1978, pp. 1303-1310.

Unite, Steve, "Beat the Chalk," *The Horseplayer Magazine*, 2000, pp. 56-58.