WORKING PAPER SERIES

WORKING PAPER NO 12, 2007

ESI

Testing Market Efficiency in a Fixed Odds Betting Market

By

Robin Jakobsson
Department of Statistics
Örebro University
robin.jakobsson@esi.oru.se

Niklas Karlsson
Department of Statistics
Örebro University
niklas.karlsson@esi.oru.se

http://www.oru.se/esi/wps
SE-701 82 Örebro
SWEDEN

ISSN 1403-0586

Testing Market Efficiency in a Fixed Odds Betting Market

Abstract

This paper tests the hypothesis of market efficiency for the fixed odds betting market of Swedish trotting head-to-head matches. The hypothesis is carried out by a Wald test within a logistic regression model. Data support rejection of semi-strong efficiency at the 5 percent level of significance, while the weak form efficiency cannot be rejected. Moreover, evaluation of the simple strategy to bet on those horses where, conditional on the estimated model, the expected profit is positive results in a profit of 7.8 percent per bet.

Key words: efficient market, betting, trotting, logit,

JEL classification: G14

1. Introduction

The growing market for sports betting has led to an increasing research into the efficiency of various sports betting markets. Studies on the three forms of efficiency; weak form, semi-strong form and strong form, set out by Fama (1970) are all represented in the literature. Ali (1977) and Snyder (1978) test the weak form of the efficient market model applied to harness racing and horse racing, respectively, in US with a pari-mutual system. In both studies, evidence on inefficiency is found. In similar studies, Asch *et al.* (1984, 1986) and Hausch *et al.* (1985) draw the same conclusion. For other systems, however, attempts have failed to reject the weak form efficiency. This is the case for Dobra *et al*. (1990), using a spread betting system and Johnson and Bruce (1992) using quoted odds. Goddard *et al*. (2004), Kuypers (2000), and Pope and Peel (1990), all studying fixed odds betting markets cannot find evidence of inefficiency in the weak form. However, as comes to papers testing for the semi-strong form, a majority find evidence of inefficiency irrespective of betting system (see Goddard *et al*., 2004; Figlewski, 1979; Gabriel and Marsden, 1990; Gandar *et al*., 1988; Kuypers, 2000).

For fixed odds betting markets there are two approaches, broadly speaking, to testing the semi-strong form of efficiency. One possibility is to use a regression-based test where the assumption of semi-strong efficiency implies certain parameter restrictions in a model. The outcome of the sports event is regressed on a function of the odds, and other predictors as well, implicit available as public information. Another possibility is to use economic efficiency tests, that is, to calculate the *ex post* returns that could have been generated by following a certain betting strategy conditional on a forecasting model. A disadvantage of the former approach is, as pointed out by Gandar *et al.* (1988), the low power of the test. However, an advantage with this approach, at least if the testing procedure is designed in a specific way, is the by-product that parameter estimates reveal the direction and magnitude of bookmakers' systematic misjudgement of certain predictors. We use the term "misjudgement" here even though the bookmakers might have incitement to skew their predictions in order to gain higher profits than if they do not skew their predictions.

In the present paper, we test the semi-strong and weak form efficiency of bookmakers' fixed odds using data on Swedish trotting head-to-head matches during a five-month period. A testing procedure is designed to test efficiency within a model. The model is also used for forecasts after eliminating nonsignificant predictors. Thus, the two approaches described above are linked in a sense that the model underlying the regression-based approach is also used to test the efficiency by the other approach.

The rest of the paper is organized as follows: In Section 2, model for the outcome of head-to-head matches is developed. The relevant parameter restrictions implied by the efficiency assumptions in the proposed model are also stated. Further, the section contains a test statistic and a brief description of a betting strategy. In Section 3, data and selected variables are presented. Section 4 contains results of estimation of the model and testing results. This section also includes results from applying the proposed betting strategy. A final section concludes the paper.

2. Model

2.1 Model Specification

We define a match as a competition between two horses, denoted by horse 1 and horse 2, selected by a bookmaker from a trotting race consisting of 12 horses. Let T_{ij} be the running time for horse i, $i = 1, 2$ in match j, $j = 1, 2, ..., n$. We assume that the intensity for T_{ij} is

$$\lambda_{ij}(t) = u_j(t)e^{\boldsymbol{\beta}'\mathbf{x}_{ij}},$$

where $u_j(t)$ is a common baseline intensity for match j. The first element of the fixed $(K \times 1)$ vector $\mathbf{x}_{ij}$ is assumed a one for $i = 1$, zero otherwise; the second element is the natural logarithm of the bookmaker's assessment of the probability of horse i winning over the other horse, a probability that can easily be derived from the fixed odds. The last $K-2$ elements are horse- and driver characteristics. The corresponding parameter vector $\boldsymbol{\beta}'$ has dimension $(1 \times K)$. It can now be shown that the probability of horse 1 winning over horse 2 in an arbitrary match j is

$$\pi(\mathbf{z}_j;\boldsymbol{\beta}) = P(T_{1j} < T_{2j} | \mathbf{z}_j;\boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}'z_j}}{1+\mathrm{e}^{\boldsymbol{\beta}'z_j}}, \qquad \text{where } \mathbf{z}_j = \mathbf{x}_{1j} - \mathbf{x}_{2j}.$$

Thus, we end up with a logistic regression model, where the probability of horse 1 winning over horse 2 in a certain match is allowed to depend on the difference in the natural logarithm of the bookmaker's probability judgement derived from the odds. The probability also depends on the differences in horse- and driver characteristics between the two horses.

To each match we can associate a dichotomous variable, d_j, taking the value one if horse 1 beats horse 2 and zero otherwise. Then, the log-likelihood function is given by

$$\ln L = \sum_{j=1}^{n} \left[d_j \ln(\pi(\mathbf{z}_j;\boldsymbol{\beta})) + (1-d_j)\ln(1-\pi(\mathbf{z}_j;\boldsymbol{\beta})) \right].$$

Maximizing the log-likelihood with respect to $\boldsymbol{\beta}$ gives us the maximum likelihood estimator $\mathbf{b}$.

2.2 Parameter Restrictions and Wald Test

It can easily be shown that, under the restrictions $\beta_1 = \beta_3 = ... = \beta_K = 0$ and $\beta_2 = 1$, the probability of horse 1 winning the match coincides with the bookmaker's probability assessment of that event. Thus, within the model specification outlined above it is possible to test the hypothesis of efficiency of semi-strong form, that is, that bookmakers take into account information on horse- and driver- characteristics available to public in an efficient way when determining the odds. We aim to test the null hypothesis

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$$

against the alternative hypothesis

$$H_1 : H_0 \text{ is not true,}$$

where $\mathbf{R} = \mathbf{I}_K$, the identity matrix of order K, and

$$\mathbf{q}' = \begin{bmatrix} 0 & 1 & 0 & 0 ......0 & 0 \end{bmatrix}$$

is a $(K \times 1)$ vector. It is straightforward to base a test on the Wald criterion:

$$W = \mathbf{d}' \left[Var(\mathbf{d}) \right]^{-1} \mathbf{d},$$

which has an asymptotically chi-squared distribution with $J = K$ degrees of freedom if the null hypothesis is true, where $\mathbf{d} = \mathbf{Rb} - \mathbf{q}$ and $Var(\mathbf{d}) = \mathbf{R}\left[Var(\mathbf{b}) \right]^{-1} \mathbf{R}'$. We note that the weak form efficiency assumption corresponds to

$$\mathbf{q} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

since the horse- and driver characteristics, in this case, are not included.

2.3 Betting Strategy

To each bet, before we know the outcome of the match, we can associate a dichotomous stochastic variable V_{ij} representing the profit of betting on horse i in match j. Conditional on certain values of the parameter vector $\boldsymbol{\beta}$ and the vector of data $\mathbf{z}_j$ we can determine the probability distribution of V_{1j}, shown in Table 1, where O_1 is the fixed odds for horse 1 and the stake is without loss of generality standardized to 1. The probability distribution of V_{2j} is obvious. A reasonable strategy is to make a bet on a horse if the model predicts a positive expected profit, that is if, for horse 1 the condition $\pi(\mathbf{z}_j;\boldsymbol{\beta}) \times O_1 - 1 > 0$ is fulfilled and, for horse 2, $[1-\pi(\mathbf{z}_j;\boldsymbol{\beta})] \times O_2 - 1 > 0$.

Table 1. Probability distribution of V_{1j}

v_{1j}	$P(V_{1j} = v_{1j})$
-1	$1-\pi(\mathbf{z}_j;\boldsymbol{\beta})$
$O_1 - 1$	$\pi(\mathbf{z}_j;\boldsymbol{\beta})$

We evaluate this strategy in Section 4 by using a cross validation approach in the following way: The data is divided into ten parts. In the first stage, the first 90 % of the observations are used for estimation and remaining 10 percent are used for prediction. In the second stage, the first 80 % of the observations and the last 10 % are used for estimation and the remaining 10 % are used for prediction. This procedure is repeated until the last 90 % of the observations are used for estimation and the first 10 % are used for prediction.

2.4 Method for evaluation of strategy

Let W_m denote the profit of the $m:th$ bet given the proposed strategy, $m = 1,..., M$, where M is the number of bets during the evaluation period. Note that, unconditional the odds and unconditional knowledge of opponents, $W_1,..., W_M$ can be regarded as identically and independently distributed random variables taking either the value -1 or a value in the interval $(0, \infty)$. Furthermore, let $E(W_m) = \mu$ and $Var(W_m) = \sigma^2$, $m = 1,..., M$. To test

$$H_0 : \mu \leq 0$$

against the alternative hypothesis

$$H_1 : \mu > 0$$

we therefore make use of the test statistic

$$Z = \frac{\overline{W} - 0}{S_W / \sqrt{M}}$$

which, under the assumptions outlined above, asymptotically follows a standard normal distribution under the null hypothesis, where $\overline{W} = \frac{1}{M}\sum_{m=1}^{M} W_m$ and

$$S_W = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(W_m - \overline{W})^2} .$$

3. Data

3.1 Data characteristics

In the present paper we utilize data from Swedish trotting races over 2140 meters with auto start (2140a); manually collected at the homepage of ATG[1]. Some of the matches, where unsatisfying information on the competing horses was available, were left out. This lack of information concerns especially foreign horses and horses younger than three years. In total, 2091 head-to-head matches with fixed odds supplied by three different bookmakers during

[1] Internet address is www.atg.com.

the time 29-11-2003 until 20-04-2004 are included in the study. These odds were collected at the homepages of the three bookmakers Expect, Nordicbet and Unibet.[2]

3.2 Variables in the data set

To test the efficiency assumptions we consider the transformation of the odds to get the bookmakers' probability judgement for a particular horse to win defined for horse 1 and 2 respectively to be

$$P_{b1} = \frac{O_2}{O_1 + O_2} \quad , \qquad P_{b2} = \frac{O_1}{O_1 + O_2}$$

where O_1 and O_2, as mentioned earlier, are the fixed odds on horse 1 and horse 2, respectively.[3] In order to get the attractive looking parameter restrictions implied by the efficiency assumptions we construct a variable, denoted by LNP_b, as the natural logarithm of P_b.

What is relevant statistical information in order to predict the results of trotting races? It is not hard to think of a number of horse characteristics that may influence the outcome of a match. As a potential indicator of the horse's shape, we incorporate $TIME$, the average time per kilometre at the latest performed race at this 2140a length, as a predictor in the model. Other possible shape indicators are *HFP* and *HSTP*, the percentage of the latest ten performed races the horse has come on first place and either second or third place, respectively. Yet two more predictors, related to the horse is taken into account, *GLS*, a dummy variable taking the value one if the horse fell into gallop in the latest start and *GP*, the proportional number of times the horse galloped in the latest ten performed races. In order to allow for a variation in the drivers' skill we consider the predictor DWP, the proportional number of times that the previous year's races the driver ended at first place. Finally, since some starting track positions are better than others are, this fact is considered by defining and incorporating the predictor STPWP, the relative frequency, measured in percent of wins for the twelve different starting track positions.

---

[2] Internet addresses are www.expect.com, www.nordicbet.com and www.unibet.com.

[3] We are aware of the fact that the interpretation of BP_1 as the bookmakers' probability judgement might be wrong. See Kuypers (2000) for an interesting

4. Results

4.1 Testing the Parameter Restrictions

This subsection presents the results of the estimation of the model as well as the results of testing the parameter restrictions set out in Section 2. Before commenting on the various parameter estimates and their significance we conclude that the observed value of the Wald statistic used to test the semi-strong form of efficiency, outlined in Section 2, is 24.60. The statistic is chi-squared distributed with, in this particular case, 9 degrees of freedom. This yields a p-value of 0.0034. Thus, we are quite confident that bookmakers do use the information on horse characteristics available to public in an efficient way when judging the probabilities of different outcomes of the matches. We do not find any support for rejecting the weak form efficiency assumption at any comfortable significance level since we get an observed value of the test statistic equal to 2.31, corresponding to a p-value of 0.31.

Table 2: Parameter estimates for the full model

Variable	Estimate	Standard error	p-value
Constant	-0.019	0.052	0.710
LNP_b	0.744	0.181	0.000
TIME	-0.216	0.052	0.000
HFP	0.000	0.003	0.960
HSTP	-0.006	0.003	0.025
GLS	-0.021	0.093	0.820
GP	-0.001	0.002	0.710
DWP	0.009	0.008	0.245
STPWP	-0.002	0.012	0.843

As can be seen from Table 2, the parameter estimate of β_1 is not significantly different from zero, a not very surprising result. This means that data does not support bookmakers systematically to misjudge the odds depending on whether the horse is labelled "1" or "2".

There is little doubt, if any, that the variable, reflecting the bookmakers' probability judgement, is of help to explain the outcome of the match. Moreover, the fact that the estimate of the corresponding parameter β_2 is less than unity is interpreted as if bookmakers tend to overestimate the probabilities for favourites to win, when controlling for horse characteristics. If so, they tend to underestimate the win probabilities for underdogs. Be aware though that, when making this interpretation, the parameter estimate 0.744 is not significantly different from one at a 5 percent significance level. The value one, of the parameter, would indicate that the bookmakers set unbiased probability judgements.

Next, turning our attention to the predictors representing horse characteristics, we note that at least two out of seven predictors seem to be relevant to incorporate in a model, whose purpose is to predict the probabilities of the outcomes of the match as accurate as possible. The support of including two more predictors in addition to the predictor reflecting the bookmakers' probability judgement is of course interesting, since it reveals a systematic misjudgement of determining the odds by the bookmakers. The negative sign of $\hat{\beta}_3$ is interpreted as the bookmakers seem to underestimate the value of a low average time per kilometre for the horse at the latest performed race as a shape indicator. In a similar way, we can interpret the negative sign of $\hat{\beta}_5$ as an overestimate of the value of a high percentage of either second or third places.

4.2 The Model used for Prediction and Evaluation of the Strategy

The estimation results in the previous subsection indicate that the bookmakers misjudge the effects of a few variables. Thus, for predicting purposes we reduce the number of explanatory variables. This reduction was performed according to the technique of backward elimination, where the drop out criterion is chosen to a p-value above 5 %. The final model for prediction is presented in Table 3. Note that the constant term is left out for reasons discussed in the previous section.

Table 3: Parameter estimates for the reduced model

Variable	Estimate	Standard error	p-value
LNPb	0.736	0.151	0.000
TIME	-0.217	0.051	0.016
HSTP	-0.006	0.002	0.000

The cross validation approach described in Section 2 yields a profit of 7.8 percent per bet, based on those 744 bets that fulfilled the necessary requirement of positive estimated expected return. The observed value of the standard normal statistic W outlined above is 1.62 yielding a p-value of 0.053. Although we have some support for the alternative hypothesis we cannot exclude the possibility that the excess return is due to chance.

5. Summary and Conclusions

The main purpose of the present paper is to test the hypothesis that bookmakers, when determining fixed odds for betting on harness racing, use information available to public in a way that coincides with the theories of an efficient market. Furthermore, if the hypothesis is rejected we evaluate the simple strategy to bet on those horses where, conditional on an estimated binary logistic regression model, the expected profit is positive. The result of the test indicates that the bookmakers do not tend to use the information efficiently. Moreover, the data supports that bookmakers misjudge the effects of at least two horse characteristics; time per kilometre for the horse at the latest performed race and the percentage of the latest ten performed races the horse has come on either second or third place. Furthermore, the strategy only to bet on those horses where the predicted expected profit is positive yields a profit of 7.8 percent per bet during the evaluating period. A minor purpose is to test the weak form efficiency. Contrary to the semi-strong form, we do not find evidence that the market is lacking of efficiency in this form.

The proposed approach in this paper to model and predict the outcome of head-to-head matches seems attractive. However, the way which bookmakers determine their odds is probably changing over time. Therefore, it might be a good idea to update the model estimation continuously by using as actual data as possible.

Furthermore, our choice of predictors is certainly not optimal. Instead of our time predictor, it would probably have been better to include a predictor like the difference in time between the actual horse and the winning horse the latest race, since the condition of the track differs between days and location. Unfortunately, we did not get access to that information at the time of collecting the data.

Finally, the approach of modelling the intensity for the running time can be extended to situations that are more general, for example to model the ordering of all 12 horses in a trotting race. Then it would be possible to identify which, if any, of the 12 competing horses that is worth betting on.

References

Ali, M. M. (1977) Probability and utility estimates for race track bettors. *Journal of Political Economy* **85:** 803-815.

Asch, P., Malkiel, B. G. and Quandt R. E. (1984) Market efficiency in race track betting. *Journal of Business* **57**: 157-175.

Asch, P., Malkiel, B. G. and Quandt R. E. (1986) Market efficiency in race track betting: further evidence and a correction. *Journal of Business* **59**: 157-160.

Dobra, J. L., Cargill, T. F. and Meyer, R. A. (1990) Efficient markets for wagers: the case of professional basketball wagering, in *Sportometrics* (Eds) B. Goff and R. D. Tollison. Texas A & M University Press, pp. 215-249.

Fama, E. F. (1970) Efficient capital markets: a review of theory and empirical work. Journal of Finance **33**: 383-423.

Figlewski, S. (1979) Subjective information and market efficiency in a betting market. *Journal of Political Economy* **87**: 75-88.

Gabriel, P. E. and Marsden, J. R. (1990) An examination of market efficiency in British racetrack betting. *Journal of Political Economy* **98**: 874-885.

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

Goddard J. and Asimakopoulos I. (2004) Forecasting Football Results and the Efficiency of fixed-odds Betting. *Journal of Forecasting* **23**: 51-66.

Hausch, D. B. and Ziemba, W. T. (1985) Transactions costs, extent of inefficiencies, entries and multiple wagers in a racetrack betting model. *Management Science* **31**: 381-394.

Johnson, J. E. V. and Bruce, A. C. (1992) Betting pattern in early and late races: an exploitable opportunity?, Discussion papers in Accounting and Management Science, University of Southampton.

Kuypers, T. (2000) Information and efficiency: an empirical study of a fixed odds betting market. *Applied Economics* **32**: 1353-1363.

Pope, P. F. and Peel, D. A. (1990) Information, prices and efficiency in a fixed-odds betting market. *Economica* **56**: 323-341.

Snyder, W. W. (1978) Horse racing:testing the efficient marketsmodel. The Journal of finance **XXXIII**: 1109-1118.