HAL
archives-ouvertes.fr

Nonlinear Modeling of European Football Scores Using Support Vector Machines

Raphael Nicholas Markellos

▶ To cite this version:

Raphael Nicholas Markellos. Nonlinear Modeling of European Football Scores Using Support Vector Machines. Applied Economics, Taylor & Francis (Routledge), 2008, 40 (01), pp.111-118. 10.1080/00036840701731546 . hal-00582282

HAL Id: hal-00582282
https://hal.archives-ouvertes.fr/hal-00582282

Submitted on 1 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

For Peer

Nonlinear Modeling of European Football Scores Using Support Vector Machines

Journal:	*Applied Economics*
Manuscript ID:	APE-07-0722
Journal Selection:	Applied Economics
Date Submitted by the Author:	05-Oct-2007
Complete List of Authors:	Markellos, Raphael; Athens University of Economics and Business, Management Science and Technology
JEL Code:	C45 - Neural Networks and Related Topics < C4 - Econometric and Statistical Methods: Special Topics < C - Mathematical and Quantitative Methods, C53 - Forecasting and Other Model Applications < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods, G14 - Information and Market Efficiency|Event Studies < G1 - General Financial Markets < G - Financial Economics
Keywords:	C45 - Neural Networks and Related Topics < C4 - Econometric and Statistical Methods: Special Topics < C - Mathematical and Quantitative Methods, C53 - Forecasting and Other Model Applications < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods, G14 - Information and Market Efficiency|Event Studies < G1 - General Financial Markets < G - Financial Economics

For Peer Review

Nonlinear Modeling of European Football Scores Using Support Vector Machines

Nikolaos Vlastakis[(i)], George Dotsis[(ii)], Raphael N. Markellos[(iii)]

For Peer Review

Abstract. This paper explores the linear and nonlinear forecastability of European football match scores using IX2 and Asian Handicap odds data from the English Premier league. To this end, we compare the performance of a Poisson count regression to that of a nonparametric Support Vector Machine (SVM) model. Our descriptive analysis of the odds and match outcomes indicate that these variables are strongly interrelated in a nonlinear fashion. An interesting finding is that the size of the Asian Handicap appears to be a significant predictor of both home and away team scores. The modeling results show that while the SVM is only marginally superior on the basis of statistical criteria, it manages to produce out-of-sample forecasts with much higher economic significance.

Keywords : Betting, Football, Forecasting, Nonlinear Models, Support Vector Machines

JEL Classification : L83, C53, C45, G14

(i) PhD Candidate, Department of Management Science and Technology, Athens University of Economics and Business, Greece.
(ii) Lecturer, Department of Accounting, Finance and Management and Essex Finance Centre, University of Essex, Colchester, CO4 3SQ, UK.
(iii) Corresponding Author: Senior Lecturer, Department of Management Science and Technology, Athens University of Economics and Business, Office 915, 47A Evelpidon Str. 113 62, Athens, Greece; Visiting Research Fellow, Centre for Research in International Economics and Finance (CIFER), Loughborough University, UK. E-mail: rmarkel@aueb.gr ; Tel. +30 210 8203671; Fax. +30 210 8828078.

1. Introduction

The usefulness of nonlinear models in forecasting financial variables and challenging market efficiency has been extensively investigated over the past twenty years (see Mills and Markellos, 1997, *inter alia*). Despite the fact that the efficiency of wagering markets has been also widely studied, most of this research has employed linear parametric models (see Sauer, 1998; Vaughan Williams, 2005; Vlastakis *et al.*, 2007). As was the case with financial markets, one could reasonably argue that the inconclusiveness of results with respect to betting market efficiency may be due to a misspecification of the models used.

The present paper extends the literature on market efficiency by evaluating the statistical and economic performance of a new class of nonparametric regression models, namely Support Vector Machines (SVMs), in forecasting the outcome of European football matches using odds information from 5 UK bookmakers. The results are compared to those obtained by a standard parametric approach based on Poisson count regression. SVMs belong to the family of neural networks which have been widely applied within the financial literature (see, for example, McNelis, 2005). Chen *et al.* (1994) were among the first to employ such techniques for predicting the results of greyhound races. They adopted a decision tree building algorithm along with a backpropagation Artificial Neural Network (ANN) and tested their performance against predictions from human specialists. They found that their techniques were able to outperform their human competitors, with the ANN achieving the best performance. Johansson and Sonstrod (2003) also used a similar ANN approach in modeling greyhound race outcomes and were able to “beat the

For Peer Review

market" and achieve positive returns. Rotshtein *et al.* (2005) employed a fuzzy knowledge base with genetic and neural tuning in order to predict the outcome of European football matches from a dataset of Finnish championship matches for the period 1994-2001. Their somewhat complex model displayed superior predictive ability on the basis of statistical criteria. Recently, Edelman (2007) adapted the methodology of SVMs for predicting the outcome of horse races. His model employed past performance data and bookmaker's odds in a two-stage approach. A "Winningness Index" forecast was first obtained from the SVM and then this was used along with bookmaker odds in a multinomial logit model in order to obtain probability forecasts. The author tested his methodology on a small sample of Australian horseracing data are reported promising results.

The rest of the paper is organized as following. The next section discusses the methodology used. Section 3 presents the empirical results, whereas the final section concludes the paper.

2. Methodology

Poisson Count Regression

As has been widely discussed in the literature, the Poisson process is a natural assumption when dealing with count data, such as the number of goals scored by a team in a football match. The density of the distribution of the number of occurrences of the event is given by

$$\Pr(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \quad y = 0,1,\ldots, \tag{1}$$

For Peer Review

where λ is the mean and variance parameter. It has been shown in sports betting literature (eg., see Dixon and Coles, 1997; Cain, Law and Peel, 2000) that the goal scoring processes of the home and away teams can be approximated by independent Poisson processes.

The Poisson regression model is derived from the Poisson distribution through the parameterization of the relation between the mean parameter λ and the regressors $\boldsymbol{x}$ (for a comprehensive description of count regression see Winkelmann, 2003). The standard method involves an exponential mean parameterization:

$$\lambda_i = e^{x_i'\beta}, \ i = 1,\ldots,n \tag{2}$$

The natural estimator for the Poisson regression model is maximum likelihood (ML) using the following cost function

$$\ln L(\beta) = \sum_{i=1}^{n} \left\{ y_i x_i'\beta - e^{x_i'\beta} - \ln y_i ! \right\} \tag{3}$$

The Poisson ML is the solution to k - the number of covariates - nonlinear equations corresponding to the first-order condition for maximum likelihood

$$\sum_{i=1}^{n} \left(y_i - e^{x_i'\beta} \right) x_i = 0 \tag{4}$$

For Peer Review

Two separate models are estimated, one for the home team score and the other for the away team score.

Support Vector Machine Regression

The SVM methodology is similar to that used for building ANNs. Given a sample of training data $\{(x_1, y_1),...,(x_n, y_n)\} \subset \mathbb{R}^m \times \mathbb{R}$, where *m* is the number of regressors, the goal in the so-called *ε*-SV regression is to find a function $f(x)$ that deviates from the targets y_i by a maximum of *ε*, and, at the same time, is as flat as possible (see Vapnik, 1995). In the case of linear functions, f takes the form

$$f(x) = \langle w, x \rangle + b \text{, with } w \in \mathbb{R}^m, b \in \mathbb{R} \tag{5}$$

Where $\langle \cdot, \cdot \rangle$ denotes the dot product in $\mathbb{R}^m$. Flatness in this context means that the goal is to find a small *w*. In order to ensure this, one needs to minimize the norm, i.e., $\|w\|^2 = \langle w, w \rangle$. This problem can be formulated as a convex optimization problem

$$\text{minimize } \tfrac{1}{2}\|w\|^2$$

$$\text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon \\ \langle w, x_i \rangle + b - y_i \le \varepsilon \end{cases} \tag{6}$$

For Peer Review

The assumption in (5) is that there exists a function f that can approximate all pairs (x_i, y_i) with ε precision. In other words, this means that the optimization problem is feasible. Since this is not always the case, we can cope with otherwise infeasible constraints of the optimization problem by introducing slack variables ξ_i, ξ_i^*. Thus, the optimization problem takes the form first introduced in Vapnik (1995)

$$\text{minimize} \quad \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right)$$

$$\text{subject to} \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{7}$$

The constant $C > 0$ determines the trade-off between flatness and the amount up to which deviations larger than ε are tolerated. In order to reach the nonlinear form of f, the dual formulation of the optimization problem in (6) is needed. To this end, some kind of dualization method is required, the most common utilizing Lagrange multipliers. This method constructs a Lagrange function from the objective function and the corresponding constraints, by introducing a dual set of variables. The dual optimization problem takes the following form

$$\text{maximize} \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right)\langle x_i, x_j \rangle \\ -\varepsilon\sum_{i=1}^{n}\left(a_i + a_i^*\right) + \sum_{i=1}^{n} y_i\left(a_i - a_i^*\right) \end{cases}$$

For Peer Review

$$\text{subject to } \sum_{i=1}^{n}\left(a_i - a_i^*\right) = 0 \text{ and } a_i, a_i^* \in [0, C] \tag{8}$$

$$\text{In this formulation, } w = \sum_{i=1}^{n}\left(a_i - a_i^*\right)x_i \text{, thus } f(x) = \sum_{i=1}^{n}\left(a_i - a_i^*\right)\langle x_i, x\rangle + b \tag{9}$$

This is the so-called *Support Vector Expansion*. Now w can be completely described as a linear combination of the training patterns x_i. For the transition to the nonlinear form of the algorithm, the best method is the so-called implicit mapping with the use of kernels. Since the SV algorithm depends only on dot products between patterns x_i, it is enough to know $k(x_i, x) := \langle \Phi(x_i), \Phi(x) \rangle$, rather than Φ explicitly, so that w is a nonlinear combination of the training patterns x_i and the optimization problem becomes

$$\text{maximize } \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{n}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right)k\left(x_i, x_j\right) \\ -\varepsilon\sum_{i=1}^{n}\left(a_i + a_i^*\right) + \sum_{i=1}^{n} y_i\left(a_i - a_i^*\right) \end{cases}$$

$$\text{subject to } \sum_{i=1}^{n}\left(a_i - a_i^*\right) = 0 \text{ and } a_i, a_i^* \in [0, C] \tag{10}$$

$$\text{where } w = \sum_{i=1}^{n}\left(a_i - a_i^*\right)\Phi(x_i) \text{ and } f(x) = \sum_{i=1}^{n}\left(a_i - a_i^*\right)k(x_i, x) + b \tag{11}$$

In our case a Gaussian kernel is employed, hence

$$k(x_i, x) = e^{-\frac{\|x_i - x\|^2}{2\sigma^2}} \tag{12}$$

For Peer Review

As is the case for the Poisson model, two separate SVM models are estimated, one for the score of each team.

3. Data and Empirical Results

Our dataset contains match results from the English Premier League for the period 13/8/2005 to 10/5/2007, a total of 750 observations. We also have the corresponding 1X2 odds from 5 online bookmakers: Gamebookers, Interwetten, Ladbrokes, Sportingbet, and William Hill, coded hereafter as A, B, C, D, and E, respectively. Data on Asian Handicap odds along with the handicap size were also available for the matches under study.

In order to gain a clearer understanding of the relationship between odds, Figure 1 presents 3-dimensional (3D) scatter plots of the odds quoted by each bookmaker for all the matches in the sample. Every point in the 3D space represents a match. The position of each point depends on the value of the odds for the three possible outcomes, whereas the color represents the actual outcome: red for home team victory, blue for draw and yellow for away team victory. A clearly nonlinear relationship appears to exist between odds and match outcomes for all bookmakers. The diagram indicates that when there is a strong favourite, the odds become a relatively better predictor of match outcomes. The distribution of points could be divided into three segments. The first contains the matches for which the home team is the favourite and is the part of the distribution that contains high values for the odds on the away team. The second contains matches for which there is no strong

For Peer Review

favourite and is the part of the distribution that is closest to the intercept of the axes. Finally, the third segment contains the matches for which the away team is the favourite and is the part of the distribution that contains high values for the odds on the home team. If we look at the first and the third segment we realize that, in each of the two, one color is dominant and it is the color that corresponds to the favourite, whereas the second segment is a more of a random mix in term of coloring or outcomes.

[Insert Figure 1 about here]

Table 1 presents the linear correlation matrix of odds for all bookmakers in the sample. As expected, odds on the same outcome between bookmakers are generally highly correlated. The same holds for the odds on different outcomes for each bookie. Asian Handicap odds Exhibit small correlation with 1X2 odds, in contrast to the handicap size, which is significantly correlated to the odds on all outcomes for most bookmakers. In terms of the correlation structure, bookmaker D variables appear to have a somewhat different behavior. As the variables in Table 1 are to be used as regressors, problems from multicollinearity are very likely to arise, certainly in the Poisson count regression. Indeed, as suggested by the graphical analysis the multicolinear relationship between the regressors is likely to be nonlinear.

[Insert Table 1 about here]

For Peer Review

When examining the predictive power of odds, the bookmaker margins are a very important metric, since the size of the margin directly affects the odds. However, it is not possible to know the actual margin without knowing the actual distribution of bets across outcomes. The standard practice in the literature is to calculate an *implied* margin by assuming an equal distribution of bets. Table 2 presents the descriptive statistics of implied margins for all bookmakers in the sample. An examination of the table reveals that bookmakers B, C, and E operate at comparable levels, whereas bookmakers A and D operate at considerably lower profit levels. This could either be an indication that the implied margin is not a good proxy for the actual margin or a sign of market distortion.

[Insert Table 2 about here]

The data was divided in two samples. Roughly 78% of the data (588 observations) was used for the estimation of the models, whereas the remaining data (162 observations) were used for out-of-sample evaluation purposes. Table 3 presents the estimation results of the two Poisson models. The method used for the estimation was stepwise regression using a 5% level of significance. As can be seen, the statistically significant regressors for the home score model are the constant, the size of the Asian Handicap, the odds for away team victory from bookmakers A and E and the margin of bookmaker C,. For the away team score, only the Asian Handicap size and the odds for away team victory from bookmaker A enter the regression as statistically significant variables. The signs of the coefficients are as expected, with

For Peer Review

the exception of the odd on away team victory from bookmaker A in the home score regression. The adjusted R^2 and log likelihood statistics reveal a reasonably good fit of the models.

[Insert Table 3 about here]

In order to investigate model misspecification two sets of tests were undertaken. As mentioned, one of the most important assumptions of the Poisson regression model is the equality of mean and variance. If this assumption is violated, then the Poisson model is deemed unsuitable for the specified application and other, less restrictive models, such as the negative binomial, should be used. To test the equality of mean and variance, we follow the specification by Wooldridge (1997). This is a regression-based overdispersion test that is carried out by regressing $e_{si}-1$ on $\hat{y}_i$, where e_{si} are the standardized residuals and $\hat{y}_i$ the fitted values for the dependent variable, from the original model. If the resulting coefficient is found statistically significant, then the model is over- or under-dispersed, depending on the sign of the coefficient.

The other misspesification test is a generalized version of Ramsey's RESET test for ML models, originally proposed by Peters (2000). This is a Likelihood Ratio (LR) test that is conducted by adding RESET variables, i.e., powers of fitted values $\hat{y}_i^j, j=1..k$, as regressors in the original model. If $l_0\left(\hat{\theta}\right)$ is the maximized log-likelihood function for the original model, and $l_a\left(\tilde{\Psi}\right)$ that for the extended model,

For Peer Review

then the LR statistic is constructed as $2*\left(l_a\left(\tilde{\Psi}\right)-l_0\left(\hat{\theta}\right)\right)$ and is asymptotically distributed as χ^2_{k-1}, where k is the order of the power of the fitted values used in the regression. In this paper, we will consider powers up to four, i.e., $k = 2,3,4$.

Table 4 summarizes the results of the misspesification tests. None of the reported statistics are significant in the 5% level, which indicates that the model is well-specified. Although the Generalised RESET (GRESET) should be sensitive also to departures from linearity, it may not be able to capture the complex nonlinear structure of the data suggested by the graphical analysis.

[Insert Table 4 about here]

For the estimation of the SVM regression model, the independent covariates that were used in each of the two regressions were the ones found significant for the respective Poisson models. This choice of regressors was made so that the information available to each model was the same and direct comparisons can be drawn. The value for σ^2 in equation (12) was set via trial-and-error to 20 using a crossvalidation in the fist sub-sample. The magnitude of this parameter is very important since it determines the flexibility of the model. A very low value may lead to overfitting the estimation sample, a common problem for nonparametric and nonlinear models.

Table 5 presents the evaluation of model performance using statistical error functions for the in-sample and out-of-sample data, respectively. Although the two

For Peer Review

modeling procedures have comparable performance, the SVM model outperforms the Poisson regression in almost all instances.

[Insert Table 5 about here]

Although the results reported in Table 5 suggest that the two models have similar forecasting performance this may not be true if other non-statistical cost functions are used. As argued by Satchell and Timmermann (1995), standard statistical error functions may not be suitable for assessment of the economic value of predictions of non-linear models. Thus, the final test of the forecasting performance of the two models will be the economic evaluation of the forecasts provided by the models, through the implementation of a betting strategy.

For the formulation of betting strategies, the data are divided again in two samples. One is used for the estimation of the models and the other for out-of-sample evaluation. Once the models are fitted the out-of-sample data are used to provide forecasts of the goals scored by each team. Then, a forecasted goal difference is calculated for every match. The goal difference is used as the variable under consideration for the formulation of the betting strategy. The strategy consists of a straightforward rule: "*If the forecasted goal difference is positive and greater than some threshold T_1, bet on the home team; if it is negative and less than some threshold T_2, bet on the away team; and, if its absolute value is less than some threshold T_x, bet on x*". Thresholds are estimated numerically so as to maximize total profits using in-sample data. The same thresholds are used on the out-of-sample data and the two models are evaluated with respect to economic performance.

For Peer Review

Table 6 summarizes the results of the betting strategies for both models. It can be easily seen that Table 6 reveals a different picture than before with respect to the comparative performance of the models. Although the Poisson model is able to outperform the SVM model on in-sample data and return higher total profits with higher expected return per bet, the SVM model greatly outperforms the Poisson model on out-of-sample data. In fact, the Poisson model is unable to produce positive profits in the out-of-sample data, whereas the SVM model returns significantly high profits in both samples. Consequently, the performance of the SVM model is more stable and the model itself more robust than the Poisson.

[Insert Table 6 about here]

4. Conclusions

This paper examined the weak form efficiency of the UK football betting market using a sample of odds from 5 bookmakers. Motivated by the financial literature and in order to asses the nonlinear forecastability of match outcomes we employed a Support Vector Machine modeling approach. This relatively new class of neural network models have been found in the literature able to capture a wide variety of nonlinear relationships. The performance of the SVMs was compared to that of a standard Poisson count regression. Our preliminary descriptive analysis demonstrates that regressors based on odds variables are likely to be strongly dependent in a nonlinear fashion not only between them but also against match outcomes. Although econometric tests suggest that the Poisson model regressions estimated via a stepwise procedure are well specified, these models have an inferior

For Peer Review

performance when compared to SVMs. More specifically, SVMs are slightly better in terms of out-of-sample statistical significance and far more superior in terms of the profits it produces when its forecasts are employed in a betting system. The presence of positive out-of-sample profits and the fact that the information incorporated in the models included only information on past odds, implies deviations from the weak-form efficient market hypothesis for the period considered.

For Peer Review

References

Cain, M., Law, D. and Peel, D. (2000) The Favourite – Longshot Bias and Market Efficiency in UK Football Betting, *Scottish Journal of Political Economy,* **47**, 25-36.

Chen, H., Buntin, P., She, L., Sutjaho, S., Sommer, C. and Neely, D. (1994) Expert Prediction, Symbolic Learning and Neural Networks: an Experiment on Greyhound Racing, *IEEE Intelligent Systems & their Applications*, **9**, 21-27.

Dixon, M.J. and Coles, S.G. (1997) Modelling Association Football Scores and Inefficiencies in the Football Betting Market, *Applied Statistics,* **46**, 265-280.

Edelman, D. (2007) Adapting Support Vector Machine Methods for Horserace Odds Prediction, *Annals of Operational Research*, **151**, 325-336.

Johanson, U. and Sonstrod, C. (2003) Neural Networks Mine for Gold at the Greyhound Racetrack, *Proceedings of the International Joint Conference on Neural Networks*, **3**, 1978-1801.

McNelis, P. D. (2005) *Neural Networks in Finance: Gaining Predictive Edge in theMarket*, Elsevier Academic Press.

Mills, T.C. and Markellos R. N. (2007) *The Econometric Modelling of Financial Time Series*, 3rd edition, Cambridge: Cambridge University Press (forthcoming).

Peters, S. (2007) On the Use of the RESET Test in Microeconometric Models, *Applied Economics Letters*, **7**, 361 – 365.

Rotshtein, A. P., Posner, M. and Ratikyanskaya, A. B. (2005) Football Predictions Based on a Fuzzy Model with Genetic and Neural Tuning, *Cybernetics and Systems Analysis*, **41**, 619-630.

Satchell, S. and Timmermann, A. (1995) An Assessment of the Economic Value of Non-linear Foreign Exchange Rate Forecasts, *Journal of Forecasting*, **14**, 477-497.

For Peer Review

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

Vapnik, V. (1995) *The Nature of Statistical Learning Theory*, Springer, New York.

Vaughan Williams, L. (ed.) (2005) *Information Efficiency in Financial and Betting Markets*, Cambridge: Cambridge University Press.

Vlastakis, N., Dotsis, G. and Markellos, R. N. (2007) How efficient is the European Football Betting Market? Evidence from arbitrage and trading strategies, *Working Paper*, Athens University of Economics and Business. Available at SSRN: http://ssrn.com/abstract=984469.

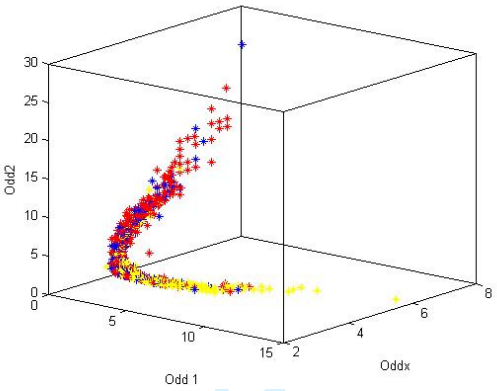
Winkelmann, R. (2003) *Econometric analysis of count data*. 4th, Berlin: Springer Verlag.

Wooldridge, J. M. (1997) A Note on the Lagrange Multiplier and *F*-statistics for Two Stage Least Squares Regressions, *Economics Letters*, **34**, 151-155.

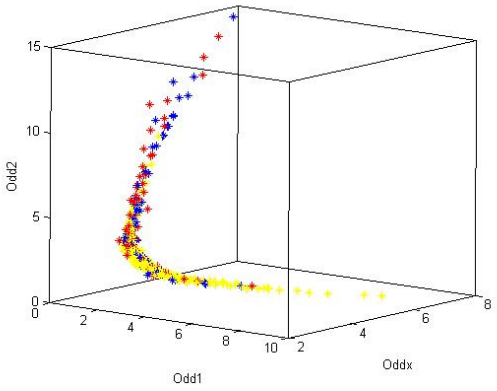
For Peer Review

Figure 1. 3D Scatter Diagrams of Odds and Outcomes

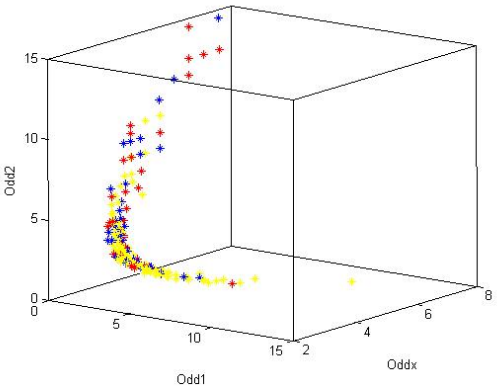
Bookmaker A



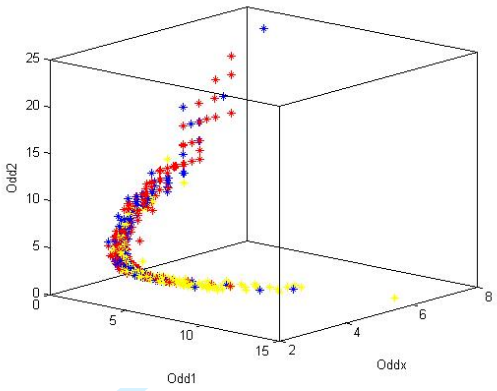
Bookmaker B



Bookmaker C



Bookmaker D



Bookmaker E

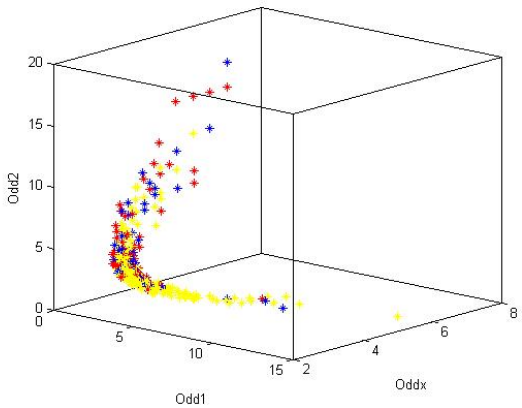


Table 1. Correlation Matrix of Odds between Bookmakers

	AHS	1(AH)	2(AH)	1(A)	1(B)	1(C)	1(D)	1(E)	X(A)	X(B)	X(C)	X(D)	X(E)	2(A)	2(B)	2(C)	2(D)	2(E)
AHS		0.06	-0.08	0.83	0.86	0.84	-0.03	0.83	-0.54	-0.53	-0.53	-0.05	-0.52	-0.87	-0.89	-0.89	0.01	-0.88
1(AH)	0.06		-0.93	-0.11	-0.11	-0.11	0.00	-0.11	-0.04	-0.03	-0.03	0.00	-0.03	0.02	0.03	0.03	0.03	0.03
2(AH)	-0.08	-0.93		0.11	0.11	0.10	-0.01	0.11	0.05	0.05	0.04	0.01	0.04	0.00	-0.01	-0.01	-0.03	0.00
1(A)	0.83	-0.11	0.11		0.99	0.99	-0.03	0.99	-0.07	-0.04	-0.07	-0.03	-0.06	-0.53	-0.57	-0.56	0.00	-0.55
1(B)	0.86	-0.11	0.11	0.99		0.98	-0.03	0.98	-0.12	-0.10	-0.13	-0.03	-0.11	-0.57	-0.62	-0.61	0.00	-0.60
1(C)	0.84	-0.11	0.10	0.99	0.98		-0.03	0.99	-0.09	-0.06	-0.09	-0.03	-0.08	-0.54	-0.58	-0.58	0.00	-0.57
1(D)	-0.03	0.00	-0.01	-0.03	-0.03	-0.03		-0.03	0.01	-0.01	0.01	0.31	0.00	0.02	0.01	0.03	0.63	0.02
1(E)	0.83	-0.11	0.11	0.99	0.98	0.99	-0.03		-0.08	-0.05	-0.08	-0.03	-0.07	-0.54	-0.57	-0.57	0.00	-0.56
X(A)	-0.54	-0.04	0.05	-0.07	-0.12	-0.09	0.01	-0.08		0.96	0.95	0.04	0.96	0.85	0.83	0.82	-0.01	0.83
X(B)	-0.53	-0.03	0.05	-0.04	-0.10	-0.06	-0.01	-0.05	0.96		0.92	0.04	0.94	0.84	0.83	0.81	-0.02	0.82
X(C)	-0.53	-0.03	0.04	-0.07	-0.13	-0.09	0.01	-0.08	0.95	0.92		0.04	0.94	0.84	0.80	0.81	-0.01	0.82
X(D)	-0.05	0.00	0.01	-0.03	-0.03	-0.03	0.31	-0.03	0.04	0.04	0.04		0.04	0.06	0.04	0.06	0.00	0.06
X(E)	-0.52	-0.03	0.04	-0.06	-0.11	-0.08	0.00	-0.07	0.96	0.94	0.94	0.04		0.83	0.81	0.80	0.00	0.81
2(A)	-0.87	0.02	0.00	-0.53	-0.57	-0.54	0.02	-0.54	0.85	0.84	0.84	0.06	0.83		0.98	0.98	-0.01	0.98
2(B)	-0.89	0.03	-0.01	-0.57	-0.62	-0.58	0.01	-0.57	0.83	0.83	0.80	0.04	0.81	0.98		0.98	-0.01	0.97
2(C)	-0.89	0.03	-0.01	-0.56	-0.61	-0.58	0.03	-0.57	0.82	0.81	0.81	0.06	0.80	0.98	0.98		0.00	0.98
2(D)	0.01	0.03	-0.03	0.00	0.00	0.00	0.63	0.00	-0.01	-0.02	-0.01	0.00	0.00	-0.01	-0.01	0.00		-0.01
2(E)	-0.88	0.03	0.00	-0.55	-0.60	-0.57	0.02	-0.56	0.83	0.82	0.82	0.06	0.81	0.98	0.97	0.98	-0.01	
Mean	-0.14	-0.08	-0.03	0.08	0.05	0.07	0.05	0.07	0.32	0.32	0.31	0.03	0.31	0.20	0.18	0.18	0.03	0.19

Odds for Home Victory, Draw and Away Victory for bookmaker A, B, C, D, E, and Asian Handicap (AH) are denoted by 1, X and 2, respectively. There are no odds for Draw in Asian Handicap. AHS is the Asian Handicap Size. The 5% and 1% two-sided critical values for the correlation coefficient are 0.062 and 0.081, respectively.

Table 2. Descriptive statistics of margins

	Bookmaker				
	A	***B***	***C***	***D***	***E***
Mean	0.0807	0.1184	0.1230	0.0838	0.1249
St. Dev.	0.0041	0.0101	0.0038	0.0725	0.0025
Min.	0.0563	0.0998	0.1113	-0.8814	0.1206
Max.	0.1012	0.1413	0.1319	0.1067	0.1313

For Peer Review

Table 3. Poisson Count ML estimation results

	Dependent Variable	
Variable	**Home Team Score**	**Away Team Score**
Constant	2.3529	-
	(2.3482)	-
AHS	-0.3586	0.5205
	(3.4301)	(8.8207)
Odd2 (A)	-0.1012	0.0202
	(2.3336)	(2.0100)
Margin (C)	-18.5739	-
	(2.2677)	-
Odd2 (E)	0.1440	-
	(2.6907)	-
Adjusted R^2	0.1511	-
Log Likelihood	-855.7344	-744.2082

Absolute values of z-statistics appear in brackets below the estimated coefficients.

Table 4. Poisson Count regression Misspecification Tests

	Overdispersion Test		Generalized RESET Test	
Dependent Variable	**Coefficient**	***t*-statistic**	**Test Variables**	**LR statistic**
Home Team Score	-0.0608	-1.6447	$\hat{y}_i^2$	1.1827
			$\hat{y}_i^2, \hat{y}_i^3$	3.7041
			$\hat{y}_i^2, \hat{y}_i^3, \hat{y}_i^4$	3.7042
Away Team Score	-0.0595	0.0489	$\hat{y}_i^2$	3.9189
			$\hat{y}_i^2, \hat{y}_i^3$	2.3012
			$\hat{y}_i^2, \hat{y}_i^3, \hat{y}_i^4$	1.3567

For Peer Review

Table 5. Statistical error functions for both models

	POISSON		SVM	
	In-sample	**Out-of-Sample**	**In-sample**	**Out-of-Sample**
Home Team Score				
MSE	1.2634	1.4999	1.2176	1.4687
MAE	0.9020	0.9690	0.8741	0.9447
RMSE	1.1240	1.2247	1.1035	1.2119
Away Team Score				
MSE	0.9416	0.9627	0.9662	0.9509
MAE	0.7667	0.7572	0.7712	0.7662
RMSE	0.9704	0.9812	0.9662	0.9751

MSE, MAE, and RMSE stand for Mean Squared Error, Mean Absolute Error, and Root Mean Squared Error, respectively.

Table 6. Economic evaluation of models

	POISSON		SVM	
	In-sample	**Out-of-Sample**	**In-sample**	**Out-of-Sample**
Number of Bets				
Bets 1	363	93	204	57
Bets X	30	9	48	15
Bets 2	99	25	67	16
All Bets	492	127	319	88
Expected Return				
Bets 1	9.87%	0.31%	13.32%	4.98%
Bets X	8.83%	7.22%	-12.29%	8.00%
Bets 2	3.65%	-7.64%	0.16%	7.56%
All Bets	8.56%	-0.76%	6.70%	5.97%
Total Profit				
Bets 1	35.84	0.29	27.17	2.84
Bets X	2.65	0.65	-5.9	1.2
Bets 2	3.61	-1.91	0.11	1.21
All Bets	42.10	-0.97	21.38	5.25

For Peer Review