International Journal of Forecasting 28 (2012) 532–542

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Simulating a basketball match with a homogeneous Markov model and forecasting the outcome

Erik Štrumbelj, Petar Vračar *

Faculty of Computer and Information Science, Tržaška 25, 1000 Ljubljana, Slovenia

A R T I C L E   I N F O

Keywords:
Sports forecasting
Probability forecasting
Monte Carlo
Simulation
National Basketball Association

A B S T R A C T

We used a possession-based Markov model to model the progression of a basketball match. The model's transition matrix was estimated directly from NBA play-by-play data and indirectly from the teams' summary statistics. We evaluated both this approach and other commonly used forecasting approaches: logit regression of the outcome, a latent strength rating method, and bookmaker odds. We found that the Markov model approach is appropriate for modelling a basketball match and produces forecasts of a quality comparable to that of other statistical approaches, while giving more insight into basketball. Consistent with previous studies, bookmaker odds were the best probabilistic forecasts.

1. Introduction

As Stekler, Sendor, and Verlander (2010) recently pointed out in their survey of sports forecasting, a large amount of effort is spent in forecasting the outcomes of sporting events. While part of this effort focuses on the properties of different types of forecasts and forecasters, the majority of such research is driven by sports betting and revolves around betting markets. The evolution of betting markets brings new challenges for forecasters. In the past, only a few different betting markets were available for a sports event. Today, bets can be made on all kinds of outcomes, ranging from the number of rebounds in a basketball match to the number of corners in a football match. Furthermore, the increasing accessibility of in-play betting has brought a demand for real-time forecasts based on the progression of the sporting event. One way of dealing with these issues in a uniform way is to construct a model that can simulate the sporting event. In this paper we take a step in this direction by exploring the use of Markov models for basketball match simulation.

1.1. Related work

Basketball has received far less attention in the scientific literature than horse racing or football. Zak, Huang, and Sigfried (1979) combined defensive and offensive elements to rank individual teams. This ranking produced the same order as the teams' actual efficiencies. However, no attempt was made to use the ranking to forecast the results of individual matches. Berri (1999) proposed a model for measuring how individual players contribute to a team's success. The sum of the contributions of all of its players was a good estimator of the actual number of wins achieved by a team. Kvam and Sokol (2006) used a Markov model to predict the winner of NCAA basketball matches. Each team was a state in the model, transition probabilities were fit using the teams' past performances, and the teams were ranked according to the model's stationary distribution. The model was better at forecasting the winner than several other statistical models and polls. Stern (1994) used the scores at the end of each quarter of 493 NBA matches to fit a Brownian motion model of the score progression. The model was good at estimating in-play win probabilities, given the current points differences. For a more comprehensive list of basketball-related forecasting references, especially those which focus on forecasting match outcomes from teams' ratings or rankings, see Stekler et al. (2010).

* Corresponding author.
E-mail addresses: erik.strumbelj@fri.uni-lj.si (E. Štrumbelj), petar.vracar@fri.uni-lj.si (P. Vračar).

0169-2070/$ – see front matter © 2011 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2011.01.004

The most relevant is the work by Shirley (2007). Shirley used a Markov model with states which corresponded to how a team obtained (retained) possession and how many points were scored during the previous transition. Shirley fit a model for each team using summary statistics and showed that the simulations based on such models were good at estimating a team's win probability against an average opponent. However, the results were in-sample and the quality of the model as a forecaster was not investigated. Shirley also discussed fitting the model using exclusively in-match data (as opposed to summary statistics) and taking into account the individual teams' strengths, but did not pursue this due to insufficient data.

1.2. Research objectives

Motivated by Shirley's work, we investigated the use of Markov models for basketball match simulation. While the long term goal is a detailed and accurate simulation of a basketball match, legitimate questions can be asked about whether a more complex model can still produce accurate forecasts for something as relevant as the match outcome. Therefore, our main objective was to determine whether such a model could be used to simulate a match between two specific teams, while still producing probabilistic forecasts of a quality comparable to that of the other statistical approaches that are most commonly used to forecast the outcome of a sporting event. We have revisited Shirley's *average home vs. average away* and *team vs. average team* models and fit them using play-by-play data from 2252 NBA basketball matches. We have also proposed an extension which takes into account the strengths of the two competing teams. This is achieved by inferring the transition probabilities from the teams' summary statistics.

We found that some related basketball forecasting studies do not include an ex-ante evaluation. Those that do, usually evaluate the forecasts using the percentage of correct predictions, which is more appropriate for evaluating binary forecasts than for evaluating probabilistic forecasts. Therefore, a secondary objective is an overview of the quality of different probabilistic forecasts for basketball. We evaluate our approach and three of the most commonly used approaches in forecasting: logit regression of the outcome, a latent strength ranking model, and bookmaker odds.

The remainder of the paper is organized as follows. In the following section we describe the Markov model and analyze its properties on NBA play-by-play data. Section 3 focuses on producing a match-specific Markov model from the teams' summary statistics. In Section 4 we describe three other well-known approaches to forecasting. The results of the empirical evaluation are given in Section 5. Finally, in Section 6 we conclude the paper and discuss some ideas for further work.

2. A possession-based Markov model

In basketball, two teams (A and B) alternate possession of the ball and try to score points. Most modern approaches to basketball analysis evaluate the match at the level of possessions (see Kubatko, Oliver, Pelton, & Rosenbaum, 2007), focusing on the number of possessions a team has and how effectively they convert them into points. Shirley (2007) proposed a Markov model with states that correspond to a team obtaining (or retaining) possession. He considered the following states: restarting the action with an inbound pass (**i**), steal or non-whistle turnover (**s**), offensive (**o**) or defensive rebound (**d**), and going to the free throw line after a shooting/bonus/technical foul (**f**).[1] These states cover all moments in a game during which the teams can obtain possession of the ball. However, such a state does not imply a change in possession. For example, a team can get two consecutive offensive rebounds, thus obtaining the ball while already having possession (that is, retaining possession). Therefore, it is possible for a team to keep possession for several consecutive states.

The model takes into account the obtainment (retainment) of possession, how it was obtained, and the number of points scored during the transition. A transition encompasses everything that happens between two states. During a transition, 0, 1, 2, or 3 points are scored. This leads us to the set of states: $\{A, B\} \times \{\mathbf{i}, \mathbf{s}, \mathbf{o}, \mathbf{d}, \mathbf{f}\} \times \{0, 1, 2, 3\}$. For example, let team B miss a shot and team A rebound, then team A miss a 2pt shot, get an offensive rebound, and dunk for 2pt while being fouled for an extra free throw. The following states and transitions describe this sequence up to the free throw line: $\cdots \rightarrow_0 \mathrm{A}\mathbf{d}0 \rightarrow_0 \mathrm{A}\mathbf{o}0 \rightarrow_2 \mathrm{A}\mathbf{f}2$. Only 30 of the 40 states are actually possible. If possession was obtained by a steal, no points could have been scored on the previous transition, and scoring 3 points in a transition can never lead to a rebound.[2] This eliminates 10 states. Table 1 shows the remaining 30 states and describes the transitions between them.

First, we estimated the transition probabilities using the frequencies from the play-by-play data available for the 2007–08 and 2008–09 seasons, separately. The result was a model for a match between the average home team and the average away team for the corresponding season — the *average home vs. average away* model. For more details about the play-by-play data and how they were transformed into states, see Table 2. The NBA play-by-play data were obtained from www.basketballgeek.com.

The transition matrix can be used to estimate the number of points scored and the win probabilities. We used a Markov chain Monte Carlo model to estimate the model's stationary distribution. From this distribution we get the number of points per transition, which, multiplied by the number of transitions, gives an estimate of the number of points scored. To estimate the win probabilities, 10 000 matches were simulated (draws were ignored). Note that some teams play fast-paced basketball, while others play at a slower pace. For example, the mean number of transitions per game for the 2007/08 (2008/09) season was 267 (264) with a standard deviation of 7.9 (8.0).

[1] Note that going to the free throw line is a state, while subsequent free throws are not considered as states. Instead, they are encompassed in the transition from the free-throw state to the next state.

[2] Scoring 2 points can lead to a rebound, for example in the situation when a player has 3 free throws, and makes the first two but misses the last one.

Table 1
Outline of the transition matrix of Shirley's Markov model for basketball. The numbers/letters and the corresponding descriptions provide more details about how a transition can occur between the row and column states.

		Team A has possession															Team B has possession														
		Ai0	Ai1	Ai2	Ai3	As0	Ao0	Ao1	Ao2	Ad 0	Ad1	Ad2	Af0	Af1	Af2	Af3	Bi0	Bi1	Bi2	Bi3	Bs0	Bo0	Bo1	Bo2	Bd0	Bd1	Bd2	Bf0	Bf1	Bf2	Bf3
Team A has possession	Ai0	0					2						4		6	7	8		A	B	C				D			F			
	Ai1	0					2						4		6	7	8		A	B	C				D			F			
	Ai2	0					2						4		6	7	8		A	B	C				D			F			
	Ai3	0					2						4		6	7	8		A	B	C				D			F			
	As0	0					2						4		6	7	8		A	B	C				D			F			
	Ao0	0					2						4		6	7	8		A	B	C				D			F			
	Ao1	0					2						4		6	7	8		A	B	C				D			F			
	Ao2	0					2						4		6	7	8		A	B	C				D			F			
	Ad0	0					2						4		6	7	8		A	B	C				D			F			
	Ad1	0					2						4		6	7	8		A	B	C				D			F			
	Ad2	0					2						4		6	7	8		A	B	C				D			F			
	Af0	0	1	1			2	3	3				5	5	5		8	9	9	B					D	E	E	G	G	G	
	Af1	0	1				2	3					5	5			8	9	9						D	E		G	G		
	Af2	0	1				2	3					5	5			8	9	9						D	E		G	G		
	Af3	0					2						5				8	9							D			G			
Team B has possession	Bi0	8		A	B	C				D			F				0					2						4		6	7
	Bi1	8		A	B	C				D			F				0					2						4		6	7
	Bi2	8		A	B	C				D			F				0					2						4		6	7
	Bi3	8		A	B	C				D			F				0					2						4		6	7
	Bs0	8		A	B	C				D			F				0					2						4		6	7
	Bo0	8		A	B	C				D			F				0					2						4		6	7
	Bo1	8		A	B	C				D			F				0					2						4		6	7
	Bo2	8		A	B	C				D			F				0					2						4		6	7
	Bd0	8		A	B	C				D			F				0					2						4		6	7
	Bd1	8		A	B	C				D			F				0					2						4		6	7
	Bd2	8		A	B	C				D			F				0					2						4		6	7
	Bf0	8	9	9	B					D	E	E	G	G	G		0	1	1			2	3	3				5	5	5	
	Bf1	8	9	9						D	E		G	G			0	1				2	3					5	5		
	Bf2	8	9	9						D	E		G	G			0	1				2	3					5	5		
	Bf3	8	9							D			G				0					2						5			

0 timeout – non-free-throw foul – defense knocks ball out of bounds
1 missed last free throw, defense tipped ball out of bounds
2 missed shot or all free throws, offensive rebound
3 missed last free throw, offensive rebound
4 fouled
5 missed last free throw, loose ball foul
6 made 2pt shot, shooting foul
7 made 3pt shot, shooting foul
8 missed shot, tipped ball out of bounds – offensive foul or violation.
9 missed last FT, tipped ball out of bounds – made last free throw
A made 2pt shot
B made 3pt shot – made 3 free throws
C steal – block – turnover without whistle
D missed shot or all free throws, defensive rebound
E missed last free throw, defensive rebound
F technical foul– offensive foul when in bonus
G missed last free throw, loose ball foul

Table 2
An excerpt from the first quarter of a match between Seattle (A) and Phoenix (B).

Play-by-play data					Parsed model states			
#	Team	Player	Event	Result	Possession	Method	Pts. on previous	State
1	PHX	Amare Stoudemire	Shot	2pt	SEA	Inbound pass	2	A**i**2
2	SEA	Damien Wilkins	Shot	Missed	-			
3	PHX	Grant Hill	Def. rebound	-	PHX	Def. rebound	0	B**d**0
4	PHX	Steve Nash	Bad pass	-	SEA	Steal	0	A**s**0
5	SEA	Kevin Durant	Shot	2pt	PHX	Inbound pass	2	B**i**2
6	PHX	Shawn Marion	Shot — 3pt	Missed	-			
7	SEA	Earl Watson	Def. rebound	-	SEA	Def. rebound	0	A**d**0
8	SEA	Damien Wilkins	Shot	2pt	PHX	Inbound pass		

Table 3
Actual mean values and *average home vs. average away* model's estimates for the 2007/08 and 2008/09 seasons.

Season	Matches	Actual			Estimated		
		Home win %	Home points	Away points	Home win %	Home points	Away points
2007–08	1132	0.606	100.9	97.1	0.598	101.0	97.5
2008–09	1122	0.609	100.4	97.0	0.607	100.9	97.5

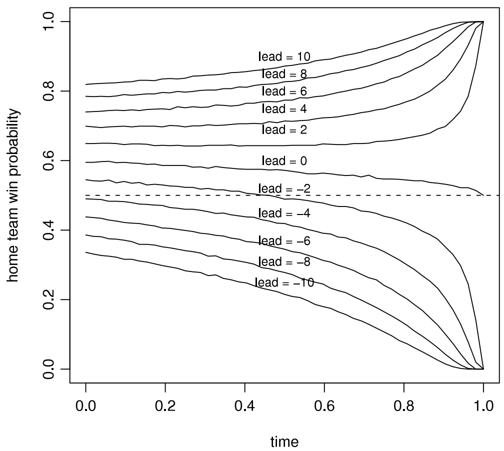


Fig. 1. Home team's win probability as a function of time and given leads of different sizes. Estimated using the 2007/08 season's *home vs. away* model.

For 2007/08 and 2008/09, the two teams with the quickest style of play were the Denver Nuggets (287) and the Golden State Warriors (285). The San Antonio Spurs had the "slowest pace" in both 2007/08 and 2008/09, with 255 and 249 transitions per game, respectively. Throughout the experiments, we use the teams' mean numbers of transitions as an estimate of the match length. When estimating for a particular team, only that team's games are used. When estimating for a match between two specific teams, the mean of the two teams' means is used.

Table 3 shows that both the 2007–08 and 2008–09 *average home vs. average away* models produce good estimates. Fig. 1 shows how the home team's win probability changes over time, given leads of different sizes. This was estimated using the 2007–08 model, and its shape is consistent with the results presented by Stern (1994). Therefore, the Markov model is capable of modelling the home team advantage and the progress of the score over time.

Shirley (2007) also fit a smaller 18-state model using the teams' summary statistics (3pt %, field goal %, rebounds, etc.) for the 2003/04 season, for each team separately — the *team vs. average opponent* model. Shirley's results suggest that the models provide a good estimate of the teams' actual win percentages ($R^2 = 0.935$). We repeated the procedure for the 2007/08 season, using the 30-state model and the play-by-play data. We estimated two models for each team, one from the team's home games and one from the team's away games. The two win probabilities estimated by these two models, weighted by the share of home (away) matches, give an estimate of the team's win percentage against an average opponent. Fig. 3(a) shows that this is a good estimate of the teams' actual win percentages. The results are similar if an out-of-sample ex-ante evaluation procedure is used (see Fig. 3(b)). We used a rolling procedure — only matches which preceded a given match were used to estimate the transition matrix for that match.

2.1. Limitations of the described Markov model approach

The teams' win percentages obtained using the Markov model and the Monte Carlo simulation are unbiased estimates of the home team win and the number of points. However, when looking at individual teams, the models overestimate the weaker teams (see Fig. 3). The same thing is observed in the results of Shirley (2007). The proposed Markov model is homogeneous — the transition probabilities do not change as the match progresses. This is a strong assumption, and those familiar with the sport will agree that it is not a valid assumption. We therefore treated each second as a separate bin and recorded the number of points scored in that particular second (see Fig. 2). Each quarter of an NBA match is 12 min or 720 s long. We pooled data from all four quarters of the 2008/09 season (the 2007/08 data give similar results). The number of points scored (and the method of scoring) remains on a similar level throughout a quarter, with the exception of approximately the first and last minutes. Those familiar with the game of basketball know how the dynamics of the match differ between time periods. In the final seconds, the

Table 4
Actual per-quarter mean values and standard deviations for both the total points scored by the two teams and the nett points.

	2007/08				2008/09			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Mean total points	50.48	49.62	49.24	49.30	49.94	49.98	49.19	49.20
(std. dev.)	8.76	8.30	8.11	8.95	8.09	8.12	7.83	8.81
Mean nett points	1.07	1.11	1.18	0.18	1.09	0.88	0.75	0.61
(std. dev.)	7.73	7.46	7.71	7.58	7.79	7.28	7.65	7.45

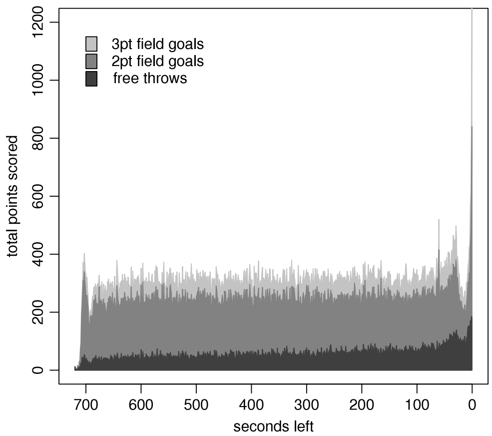
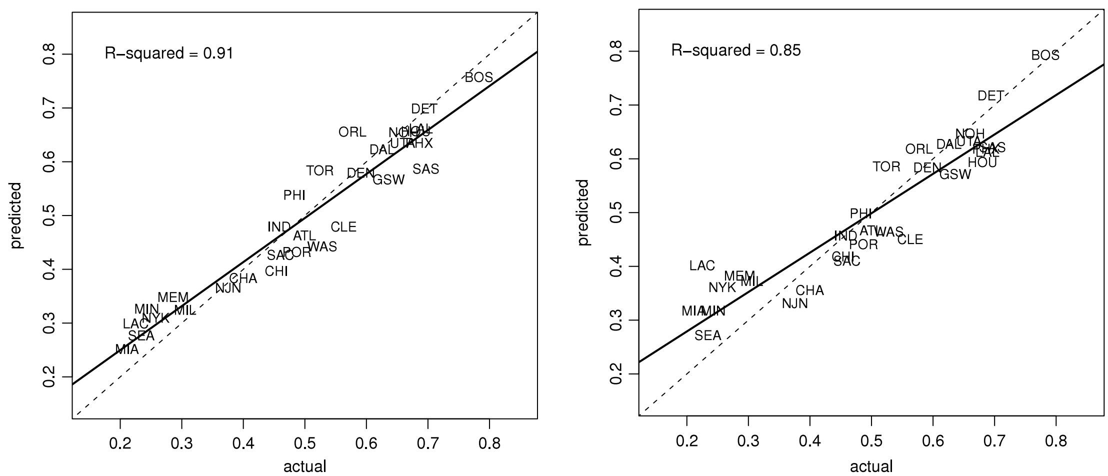


Fig. 2. Actual total number of points scored for each second-long slice of a quarter. We pooled data from all four quarters of each available match from the 2008/09 season.

teams hold up the ball and attempt to score just before time runs out to deny their opponents another possession, etc. Note also that more points are scored from free throws as a quarter progresses, because more bonus foul situations arise as the teams accumulate fouls.

Differences also exist between quarters. Table 4 shows the mean total number of points scored and the mean difference between the numbers of points scored by the teams (nett points), for each quarter. We tested the significance of the differences between quarters with a repeated measures analysis of the variance, and obtained the following p-values: for the mean total points, 0.001 and 0.006 for 2007/08 and 2008/09, respectively; for the nett points difference, 0.011 and 0.112 for 2007/08 and 2008/09, respectively. These results support the alternative hypotheses that more points are scored in some quarters than in others, and that the home team gains more (relative to the away team). Compared to the total number of points scored in a game, the differences between rgw quarters are small. Therefore, assuming homogeneity in terms of the number of points scored in a quarter does not have a strong impact on its practical use. Using all available matches, we also performed a two-way analysis of variance, using the quarter (4 categories) and minute (12 categories) as factors, and the number of points scored as the target variable. Both factors, quarter ($Pr(>|F|) = 3.1 \times 10^{-5}$) and minute ($Pr(>|F|) < 10^{-15}$), as well as the interaction quarter $\times$ minute ($Pr(>|F|) < 10^{-15}$), were found to be statistically significant sources of variance. We performed post-hoc pairwise tests with Tukey corrections to produce Fig. 4, which can be used to compare any pair of minutes with respect to the number of points scored by the two teams. Most parts of the game are similar, with the expected exceptions of the first minute of a quarter, the last minute of a quarter, and the second part of the fourth quarter, which is different from the first and the beginning of the second quarter.

Teams often deliberately play below their actual strengths. When the lead is so large that the winner is already decided, teams tend to conserve their strength and/or use substitute players. Therefore, stronger teams



(a) In-sample results.

(b) Out-of-sample results.

Fig. 3. Actual vs. estimated win percentages for the 2007/08 season.

Table 5
Best-fit linear models for the relationship between the mean nett points across the first three quarters and the mean nett points in the fourth quarter of a basketball match.

	2007/08				2008/09			
Coefficient	Estimate	Std. err.	t-value	p-value	Estimate	Std. err.	t-value	p-value
Intercept	+0.426	0.232	−1.839	0.66×10^{-1}	−0.913	0.224	−4.080	0.48×10^{-4}
Slope	−0.219	0.053	−4.116	0.41×10^{-4}	−0.331	0.052	−6.319	0.38×10^{-9}

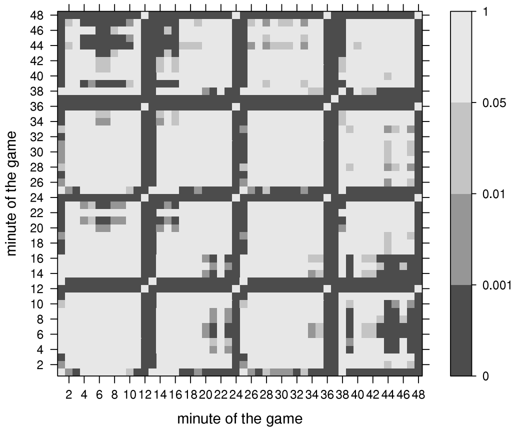


Fig. 4. Adjusted p-values for the differences in the means of the number of points scored for each pair of game minutes. The final minutes of each quarter are significantly different to all other minutes in a game.

tend to win more than their mean nett points production would suggest, because they adjust their points production towards optimizing the number of wins, not points — they play better during the decisive moments of a game and worse otherwise. This would explain the bias in the model's forecasts. To provide empirical evidence, we recorded the mean nett points over the first three quarters and in the fourth quarter. The linear models of the relationship between the two variables (see Table 5) reveal that they are negatively correlated. Therefore, on average, the larger the lead at the start of the fourth quarter, the worse a team performs in the fourth quarter. Post-hoc calibration can be used to adjust the model's forecasts for this systematic bias. However, the elegant solution of improving the underlying simulation would require a non-homogeneous model which took into account (at least) the match time and points difference.

3. Modelling the relationship between team statistics and state transitions

Within a single season, we are limited to only a few games for any given pair of teams A and B. Therefore, we do not have enough data available to enable us to estimate an opponent-specific transition matrix directly from transition frequencies; other matches have to be considered as well. First, we used the play-by-play data to model the relationship between the teams' summary statistics and transition frequencies. Models of this relationship are then used to estimate the transition matrix for any pair of A and B, given the two teams' summary statistics.

Each of the 30 rows of the Markov model (see Table 1) describes a finite set of valid transitions from the row-state to a column-state. For each row we have an exhaustive set of outcomes. We assume that the probability distribution of these outcomes depends on the shooting, rebounding, and other basketball-related skills of the two competing teams. These are most commonly described using one or more summary statistics, such as the field goal percentage.

We repeated the following procedure for each row separately for a total of 30 multinomial logit models, each describing the relationship between the teams' summary statistics and one row of the transition matrix. Let $\mathcal{G}$ be a sample of matches. First, we transform the play-by-play data into a sample that is appropriate for inferring the relationship between the teams' summary statistics and the transition probabilities. We use the first row of the transition matrix to illustrate the process. This row corresponds to state A**i**0 — the home team has an inbound pass after 0 points were scored on previous transition. Suppose that we use n different summary statistics to describe each team. Let $X_1(x), X_2(x), \ldots, X_n(x)$ and $X_{n+1}(x), X_{n+2}(x), \ldots, X_{2n}(x)$ be the values of the summary statistics of the home and away teams in match $x \in \mathcal{G}$, respectively. For each team, we compute its summary statistics across all of the team's matches in $\mathcal{G}$. Suppose that in match x there are $m(x)$ transitions from A**i**0 to another state Y. Let $\mathcal{Y}(x) = \{Y_1(x), Y_2(x), \ldots, Y_{m(x)}(x)\}$ be the set of all such states. We construct the new sample from all elements $\langle X_1(x), X_2(x), \ldots, X_{2n}(x); Y_i(x)\rangle$, $i = 1, \ldots, m(x)$, for each match $x \in \mathcal{G}$. The sample has as many elements as there are transitions from state A**i**0 in $\mathcal{G}$. The transitions from a state cannot be ordered in a meaningful way, so we used multinomial logistic regression to model the relationship between the summary statistics and the nominal dependent variable (set of valid transitions).

In our study, we used the summary statistics shown in Table 6, which are based on the four factors. Those considered the most important in determining a basketball team's success are: shooting, turnovers, rebounding, and free throws (see Kubatko et al., 2007; Oliver, 2004).

We can use the multinomial logit row models to inspect how the properties of the two teams affect the transitional probabilities. Fig. 5 shows the marginal effect of the home team's offensive rebound ratio (ORBr) on the transitions from state A**o**0. This state corresponds to an offensive rebound by team A after either a "missed shot" or "missed all free throws" (less common). The most probable transition is to state B**i**2 (made 2pt shot), but ORBr does not have a visible effect on this probability. However, ORBr does have an effect on the two most probable states that can follow from a missed shot (B**d**(0): away team defensive rebound, or A**o**0: back to the same state). As expected, a team with a higher ORBr is more likely to get an offensive

Table 6
Summary statistics for the multinomial logit row models. The base statistics are: field goals made (*FG*), field goals attempted (*FGA*), three point shots made (*3P*), free throws made (*FT*), free throws attempted (*FTA*), turnovers (*TOV*), offensive rebounds (*ORB*), and defensive rebounds (*DRB*). The prefix *o* denotes that the average of the team's opponents is used.

Effective field goal %	**Turnover ratio**	**Offensive rebound ratio**	**Defensive rebound ratio**
$EFG\% = \frac{FG+\frac{1}{2}3P}{FGA}$	$TOVr = \frac{TOV}{FGA+TOV+0.44*FTA}$	$ORBr = \frac{ORB}{ORB+oDRB}$	$DRB = \frac{DRB}{DRB+oORB}$
Free throw factor	**Opponents' EFG%**	**Opponents' TOV**	**Opponents' FTF**
$FTF\% = \frac{FT}{FTA}$	$oEFG\% = \frac{oFG+\frac{1}{2}o3P}{oFGA}$	$oTOVr = \frac{oTOV}{oFGA+oTOV+0.44*oFTA}$	$oFTF = \frac{oFT}{oFTA}$

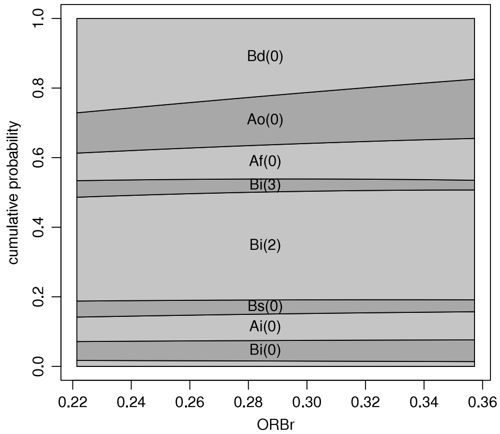


Fig. 5. The marginal effect of the home team's offensive rebound ratio on the transition probabilities from state A**o**0 (offensive rebound after 0 points were scored).

rebound. Consequently, their opponents are less likely to get a defensive rebound. It is also more likely that a team will get fouled after an offensive rebound if the team has a high offensive rebound ratio.

We conclude this section with examples of how a transition matrix for a specific game reflects the qualities of the competing teams. The first example includes two games with different home teams, the New Orleans Hornets and the Utah Jazz, versus the same away team, the Los Angeles Clippers. The two home teams were selected because of how aggressively they play (UTA: a total of 1970 personal fouls in the season, more than any other team; NOH: 1531 fouls, fewer than any other team). Table 7 shows the transition probabilities from state B**i**0 (LAC have the ball after an inbound pass with no points scored on previous possession) for these two games. NOH are more likely to allow a shot (inbound passes with points scored and defensive rebounds are more likely). On the other hand, UTA are more likely to foul, draw a foul or steal the ball (A**s**0, and B**f**/A**f** states).

The second example includes the Phoenix Suns and the Los Angeles Clippers in away games versus the Los Angeles Lakers. The two away teams were selected according to their shooting prowess (PHX were the top team with a 0.5 field goal percentage and a 0.39 3pt percentage, while LAC had only a 0.44 field goal percentage and a 0.32 3pt percentage). Table 7 shows the transition probabilities from state B**s**0 (PHX or LAC stole or blocked the ball or there was a no-whistle turnover by the Lakers). PHX is more likely to score 2 or 3 points (LAL inbound passes with points scored are more likely). On the other hand, LAC are less likely to score a two or three point shot but more likely to get an offensive rebound. The differences between the first and second examples are also interesting. Teams are more likely to score when in transition from a steal (as opposed to an inbound pass), but, on the other hand, it is less likely that the ball will be stolen back or knocked out of bounds by the defending team. This is all in accordance with basketball matches. Note that the transition to B**f**3 (made a three point shot while fouled) is very rare, regardless of the initial state.

The final example involves two teams: Dallas Mavericks, with a 0.81 free throw percentage, and Charlotte Bobcats, with a 0.71 free throw percentage (see Table 8). As a consequence, state A**f**0 is more likely to result in an inbound pass with 2 points scored when Dallas are on the free throw line (compared to CHA). When CHA are on the free throw line, the transition to an inbound pass with 1 point or a defensive rebound is more likely, because more throws are missed. Inbound passes with 2(1) points scored and defensive rebounds with 1(0) points scored cover more than 90% of all transitions from the free throw state. These transitions describe normal free throws with 2 attempts (shooting fouls), during which the last throw is either made or missed, in which case the defensive team usually gets the rebound.

4. Other methods of forecasting basketball match outcomes

The conventional approach for probabilistic forecasts of a binary event is logit (or probit) regression of the outcome. Unless otherwise noted, whenever we discuss a logit model, we are referring to using logistic regression to infer the home team win probability using a set of predictor variables, as described in this section (as opposed to multinomial logit, which refers to the row models used in the procedure we described in the previous section). Given elements $\langle X_1(x), X_2(x), \ldots, X_n(x); outcome(x)\rangle$, for each match $x \in \mathcal{G}$, we can infer the outcome directly from our summary statistics. Table 9 describes the logit regression of the home win outcome for the 2008/09 season.

We used the same summary statistics as for the multinomial logit row models in the previous section, because we wanted to observe the difference in forecasting quality between a straightforward logit regression of the outcome and indirect forecasting using multinomial logit row models (to estimate a transitional matrix for each pair of teams) and simulation. Throughout the experiments, we found that the summary statistics based on the four factors (see Table 6) were better predictors of match outcomes than basic summary statistics (shot percentages,

Table 7
Transitions probabilities from states B**i**0 or B**s**0 for games NOH vs LAC (15. 4. 2008), UTA vs LAC (28. 3. 2008), LAL vs PHX (17. 1. 2008), and LAL vs LAC (7. 3. 2008). The team listed first is the home team.

Game	State	A**i**0	A**i**2	A**i**3	A**s**0	A**d**0	A**f**0	B**i**0	B**o**0	B**f**0	B**f**2	B**f**3
A-NOH vs. B-LAC	B**i**0	0.104	0.218	0.040	0.056	0.272	0.026	0.077	0.114	0.085	0.008	0.000
A-UTA vs. B-LAC	B**i**0	0.114	0.189	0.034	0.072	0.252	0.034	0.080	0.110	0.101	0.016	0.000
A-LAL vs. B-PHX	B**s**0	0.053	0.301	0.064	0.039	0.273	0.001	0.072	0.111	0.080	0.007	0.000
A-LAL vs. B-LAC	B**s**0	0.056	0.285	0.025	0.033	0.250	0.000	0.083	0.160	0.091	0.017	0.000

Table 8
Transition probabilities from state A**f**0 for games DAL vs. GSW (2. 4. 2008) and CHA vs. GSW (5. 3. 2008). The team listed first is the home team. Note that the transitions to states A**o**2, B**d**2, and B**f**0-2 are not listed, to conserve space. The transitional probabilities for these states are all smaller than 0.001.

Game	State	A**i**0	A**i**1	A**i**2	A**o**0	A**o**1	A**f**0	A**f**1	A**f**2	B**i**0	B**i**1	B**i**2	B**i**3	B**d**0	B**d**1
A-DAL vs. B-GSW	A**f**0	0.001	0.011	0.001	0.009	0.020	0.004	0.033	0.000	0.001	0.166	0.572	0.016	0.044	0.118
A-CHA vs. B-GSW	A**f**0	0.000	0.010	0.003	0.009	0.021	0.003	0.023	0.000	0.001	0.175	0.540	0.025	0.061	0.127

Table 9
Coefficients and standard errors for the logit regression of a home team win. The model was inferred from all available matches from the 2008/09 season. A statistical summary of the predictor (see Table 6) variables. The suffixes h and a are added to denote the home and away teams.

	Logit model			Summary			
	Coefficient	Standard error	Pr(>|z|)	μ	min	max	σ
Intercept	4.648	8.866	0.600	–	–	–	–
$\mathrm{EFG\%}_h$	23.315	4.832	<0.001	0.505	0.478	0.563	0.020
TOVr_h	−28.923	9.144	0.002	0.127	0.111	0.149	0.009
ORBr_h	5.687	3.936	0.149	0.278	0.223	0.359	0.025
DRBr_h	5.444	4.212	0.196	0.743	0.680	0.792	0.021
FTF_h	5.938	3.905	0.128	0.243	0.185	0.315	0.026
$\mathrm{oEFG\%}_h$	−25.490	4.961	<0.001	0.494	0.453	0.534	0.020
oTOVr_h	19.198	8.961	0.032	0.132	0.115	0.154	0.011
oFTF_h	−5.012	3.275	0.130	0.229	0.174	0.308	0.030
$\mathrm{EFG\%}_a$	−25.911	5.810	<0.001	0.494	0.469	0.534	0.017
TOVr_a	29.321	8.235	<0.001	0.132	0.108	0.157	0.012
ORBr_a	−6.693	4.169	0.108	0.258	0.212	0.300	0.023
DRBr_a	−7.642	4.723	0.106	0.722	0.679	0.764	0.023
FTF_a	−5.096	5.292	0.336	0.228	0.182	0.273	0.022
$\mathrm{oEFG\%}_a$	20.926	5.644	<0.001	0.505	0.454	0.538	0.019
oTOVr_a	−18.742	9.867	0.057	0.127	0.109	0.150	0.010
oFTF_a	4.617	3.299	0.162	0.244	0.197	0.312	0.028

offensive/defensive rebounds per game, assists per game, fouls per game, etc.). For example, the quadratic losses for ex-ante forecasts for the 2008/09 season were 0.1977 (logit regression using the four-factors summary statistics for predictor variables) and 0.2098 (logit regression using basic summary statistics). The difference between the two was significant (p-value $\approx$ 0). Note that the quadratic loss function (also known as the Brier score) is defined as $(p_i - o_i)^2$, where p_i and o_i are the predicted probability of a home win and the actual outcome for the home team (1 = won, 0 = lost), respectively.

4.1. Latent strength modelling

Team (player) ratings or rankings are often used for forecasting the outcome of a sports event. The rating is used to describe a team's latent strength, and the ratings of a given pair of teams are used to determine the outcome of a match between the two teams. Before any matches have been played, the ratings are set to a predetermined value, which is often the same for all teams. As matches are played, the ratings are updated, depending on the teams' performances. A well-known method for rating a player's skill level is the Elo rating system. Suppose that two players (or teams) have Elo ratings of R_A and R_B, respectively. Then the expected score (in our case, win probability) for player A is

$$\Pr(\text{A wins}) = \frac{1}{1+10^{(R_A-R_B)/400}} = \frac{10^{R_A/400}}{10^{R_A/400}+10^{R_B/400}}. \quad (1)$$

Given the outcome of the match, we adjust the ratings by adding $\Delta R = K \cdot$ (actual outcome for A $-$ Pr(A wins)) to R_A and subtracting ΔR from R_B, where K is a rate-of-change constant. Note that Eq. (1) does not take into account the home team advantage. For example, observe the match between home team A and away team B, where the two teams have similar ratings. The increase in rating for team A, if they win at home, will be the same as the increase in rating for team B, if they win an away match. This is unfair, because team A had the advantage of playing at home. We compensate for this by subtracting a constant from the away team's rating

$$\Pr(\text{A wins}) = \frac{10^{R_A/400}}{10^{R_A/400}+10^{(R_B-C)/400}}, \quad (2)$$

effectively handicapping the away team.

The Elo rating system was originally developed for rating chess players, where there is no need to take the winning margin into account. However, in basketball, it is reasonable to take the winning margin into account, rewarding bigger wins more.[3] Hvattum and Arntzen (2010) used an extension of the Elo system, where the value of K is set according to the winning margin, $K = K_0(1 + \delta)^{\gamma}$, where δ is the absolute winning margin and $K_0 > 0$ and $\gamma > 0$ are parameters. Hvattum and Arntzen (2010) used this approach for ranking football teams. However, it can easily be adapted for any sport. All that is required is a proper calibration of the parameters K_0 and γ. Note that sometimes parameters other than 10 and 400 are used in Eqs. (1) and (2). The purpose of these two parameters is to set the scale of the ratings. Parameters K and C require more attention. We used the following parameter values for the basic ELO model and ELOδ: $K = 24.0$, $K_0 = 2$ and $\gamma = 1$. The home team advantage constant was set to $C = 64$. To facilitate the out-of-sample evaluation, we did not use data from 2007/08 or 2008/09 in the calibration. The chosen parameter values minimized the squared loss for the 2006/07 season. We rounded the values to simplify presentation, without significantly worsening the performance. The match scores which were used to perform the calibration were obtained from www.basketball-reference.com. References to several more complex latent-strength modelling approaches to sports forecasting, including that of Glickman and Stern (1998), are given by Stekler et al. (2010). Improvements might also be achieved by using the teams' ELO ratings (or some other measure of the team quality) and fitting a Bradley-Terry model (Bradley & Terry, 1952).

4.2. Bookmaker odds as forecasts

Bookmakers' odds are probabilistic forecasts. For example, let d_A and d_B be the quoted *decimal* odds for a match between teams A and B. Then $p_A = \frac{1}{d_A}$ and $p_B = \frac{1}{d_B}$ are win probability estimates. Due to bookmakers' profit margins, $\frac{1}{d_A}$ and $\frac{1}{d_B}$ can sum to more than 1, and thus it is common to use the normalization $p'_A = \frac{p_A}{p_A+p_B}$ and $p'_B = \frac{p_B}{p_A+p_B}$.

We used odds from the world's largest betting exchange: Betfair. Unlike regular bookmakers, betting exchanges do not offer odds directly, but simply facilitate peer-to-peer betting. Users can both back a selection (bet on an event occurring) or lay a bet (bet against an event occurring) at any odds they choose. Matching back and lay bets are matched by the system. The odds change over time, so we used the odds of the first matched bet (Betfair$_0$) and the odds most frequently matched during the 5 min before the start of the match (Betfair$_1$). These historical odds data were obtained from data.betfair.com.[4]

Bookmaker odds are considered to be the overall best source for probabilistic forecasts of sporting events. However, recent results by Franck, Verbeek, and Nuesch (2010) and Smith, Patton, and Williams (2010) suggest that betting exchange odds are better forecasts than regular bookmaker odds. However, these results are limited to horse racing and football. To the best of our knowledge, no empirical evidence exists for basketball. To test this hypothesis, we compared forecasts from 11 bookmakers and Betfair across a set of 692 NBA matches from 2008/09, such that all bookmakers' forecasts were available for each match. The selected bookmakers were 5Dimes (0.1859, 6.24), Bet365 (0.1856, 6.68), BetCRIS (0.1857, 6.76), BetUS (0.1862, 7.05), Betway (0.1866, 7.20), Canbet (0.1858, 7.01), Expekt (0.1859, 6.70), Jetbull (0.1868, 6.64), Paddy Power (0.1864, 6.37), Sports Interaction (0.1858, 6.22), Unibet (0.1866, 5.60), and Betfair (0.1855, 5.55). The numbers in parentheses represent the mean quadratic loss and mean rank (with respect to the mean quadratic loss), both across all 692 selected matches. We found no significant difference between the bookmakers' prediction qualities (p-value $\approx$1). By applying the square root to the mean quadratic loss of the best model (Betfair, 0.4307) and the worst model (Betway, 0.4320), we express them in the same unit as the predictions, thus simplifying comparison and interpretation. The difference in quality between the best and worst bookmakers is approximately one tenth of a percent. However, Betfair has the lowest mean rank. That is, the Betfair odds are, on average, closest to the actual outcome. Friedman's non-parametric test revealed that the differences in rank are significant (p-value ≈ 0). We used Nemenyi's test for post-hoc comparison. The critical difference in rank at the 0.01 significance level was 0.74. Therefore, the Betfair loss ranks significantly lower (better) than those of eight other bookmakers.

We also tested the distribution of the difference in loss between Betfair$_0$ odds and Betfair$_1$ odds. Using a pairwise t-test, we rejected the zero mean hypothesis and accepted the alternative hypothesis that the mean is greater than 0 (p-value $\approx$ 0). Therefore, the odds from just before the start of the match are significantly better probabilistic forecasts than the odds from the start of the betting period. This is consistent with the findings of Gandar et al., who showed that opening lines set by bookmakers (in our case, the market) are less accurate than the closing lines established by the market (see Gandar, Dare, Brown, & Zuber, 1998; Gandar, Zuber, & Dare, 2000).[5]

5. Empirical evaluation

We compared the following sources of forecasts: (1) the home team vs. average opponent model from Section 2 (M_{AvsAVG}), (2) the Markov model approach (using multinomial logit row models) described in Section 3 (M_{ML}), (3) Betfair$_1$ odds, (4) the Elo rating models (see Section 4.1), and (5) a logit regression model using the

[3] Some rating systems, for example Sagarin ratings, also adjust for blowouts, so that a very large margin in a single game does not have a proportionately large impact on the team's rating.

[4] The data are free of charge, but some terms apply. One must be a registered Betfair user with a certain (symbolic) number of loyalty points.

[5] Stekler et al. (2010) suggest that this indicates that the market (closing line) is more accurate than the bookmakers (opening line). We should interpret such indirect evidence with caution. Opening lines could also be worse because more information becomes available after opening lines have been set. Without all of the participants' forecasts at different points in time, we cannot distinguish between the two possibilities.

Table 10
Empirical evaluation results. The accuracy is the proportion of correct predictions.

	2007/08			2008/09		
	In-sample	Out-of-sample		In-sample	Out-of-sample	
	Loss	Loss	Accuracy	Loss	Loss	Accuracy
Betfair_1	–	0.1900	0.7042	–	0.1867	0.7106
$\frac{54}{100}\text{ELO}\delta + \frac{46}{100}M_{\text{ML}}$	–	0.1978	0.6972	–	0.1944	0.7051
ELOδ	–	0.2009	0.6835	–	0.1998	0.6871
Logit	0.1800	0.2022	0.6896	0.1765	0.1977	0.7053
M_{ML}	0.1890	0.2048	0.6872	0.1865	0.1984	0.7069
ELO	–	0.2065	0.6824	–	0.2011	0.6717
M_{AvsAVG}	0.2003	0.2199	0.6698	0.1943	0.2118	0.6853
Pr(Homewin) $= 0.6$	–	0.2373	0.6144	–	0.2371	0.6085
		$N = 957$			$N = 947$	

Table 11
Adjusted p-values for pairwise tests of the equality of means. The alternative hypothesis is that the row (model) has a lower mean quadratic loss than the column (model).

		8	7	6	5	4	3	2	1
1	Betfair_1	<0.001	<0.001	<0.001	<0.001	<0.001	0.003	0.149	–
2	$\frac{54}{100}ELO\delta + \frac{46}{100}M_{\text{ML}}$	<0.001	<0.001	0.294	0.401	0.674	0.748	–	
3	ELOδ	<0.001	<0.001	0.966	0.988	1.000	–		
4	Logit	<0.001	<0.001	0.983	0.995	–			
5	M_{ML}	<0.001	0.008	0.998	–				
6	ELO	<0.001	0.016	–					
7	M_{AvsAVG}	<0.001	–						
8	Pr(Homewin) $= 0.6$	–							

same summary statistics for predictor variables as M_{ML} (see Section 4). We also included the results of always forecasting a 0.6 home win probability and a linear opinion pool with ELOδ and M_{ML}. We included the last as an example of the usefulness of combining the forecasts of two structurally different (unbiased) models. Note that the optimization of any such linear opinion pool which included Betfair odds always resulted in the Betfair odds being weighted approximately 1. The bookmaker odds forecasts could not be improved on using a linear opinion pool. We did not investigate other types of forecast combinations.

We selected the weights (0.54 for ELOδ, 0.46 for M_{ML}) that minimized the quadratic loss on the 2006/07 data, which were reserved for parameter selection and are not used in the evaluation. In the out-of-sample procedure, we started with a sample of the first five home and first five away matches for each team (we did not include these matches in the evaluation). A rolling procedure was then used to forecast the remaining matches in a season — after forecasting and evaluating the forecasts for the next day's matches, we added the matches to the sample and rebuilt all models using the current sample of matches. This procedure was repeated until the end of the season.

We evaluated each season separately. We evaluated the models' forecasts using the quadratic loss function. We also measured the percentage of correct predictions when the forecasts were treated as being binary. For each basketball game, we had all of the models' losses (that is, a complete block design). We viewed individual games as subjects and forecasting models as treatments. We were interested in seeing whether there exist differences in mean loss between forecasting models. The equivalence of means was then tested using a one-way repeated measures procedure. We used a post-hoc Tukey test correction to account for multiple pairwise comparisons of means. Table 10 shows the results of the comparison and Table 11 shows the p-values of the significance tests. The home win forecasts and M_{AvsAVG} are significantly worse than all of the other forecasts. The Betfair_1 odds are better than all of the other forecasts. The differences in quality between the remaining statistical models are not significant. The in-sample results for some of the models illustrate how misleading their quality can be and emphasize the importance of evaluating ex-ante forecasts.

All of the statistical approaches were unbiased estimators of the win percentage, but, when observing individual teams, they exhibited a bias towards weaker teams (see Table 12). This can be explained by the fact that the statistical models do not take into account the amount of effort that was invested in a match (see Section 2.1). In other words, the summary statistics (and/or transition frequencies) are not completely accurate representations of a team's actual strength. The three statistical approaches (logit regression, ELO rating, and the proposed model) exhibit different levels of bias, but produce forecasts of the same quality. Therefore, each model represents a different bias-variance trade-off.

6. Conclusion

Using summary statistics to estimate Shirley's Markov model for basketball produced a model for a match between two specific teams. The model was be used to simulate the match and produce outcome forecasts of a quality comparable to that of other statistical approaches,

Table 12
Linear models of the relationship between the actual win percentage and the mean predicted win percentage of a team for both seasons ($N = 957 + 947$). All of the forecasters exhibit a statistically significant bias towards the weaker teams.

	M_{ML}		$\frac{54}{100}ELO\delta + \frac{46}{100}M_{ML}$		ELOδ		Logit	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Estimate	0.170	0.657	0.131	0.736	0.092	0.815	0.015	0.967
Std. Err.	0.020	0.039	0.015	0.029	0.014	0.026	0.026	0.048
t-value	8.354	16.694	8.844	25.715	6.731	30.894	0.586	19.448
$\Pr(>\lvert t\rvert)$	10^{-9}	$<10^{-16}$	10^{-9}	$<10^{-16}$	10^{-7}	$<10^{-16}$	0.563	$<10^{-16}$

while giving more insights into basketball. Due to its homogeneity, the model is still limited with respect to what it can simulate, and a non-homogeneous model is required to deal with the issues. As far as basketball match simulation is concerned, more work has to be done, with an emphasis on making the transitional probabilities conditional on the point spread and the game time.

Statistical models and bookmaker odds were better forecasters than forecasts that took into account only the home team advantage or only the strength of one team. The differences in quality between the proposed multinomial logit Markov model approach and the ELO rating models were not statistically significant. Bookmaker odds were the best source of probabilistic forecasts for NBA basketball matches, which is consistent with the results of previous studies. Recent results from horse racing and tennis suggest that betting exchange odds are better sources of information than regular bookmaker odds. However, we could not confirm that this is true for basketball forecasts.

Acknowledgements

The authors would like to thank Prof. Rob J Hyndman and the two anonymous reviewers for their detailed reviews and constructive comments on how to improve the paper.

References

Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics*, *20*, 411–427.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*, 324–345.

Franck, E., Verbeek, E., & Nuesch, S. (2010). Prediction accuracy of different market structures — bookmakers versus a betting exchange. *International Journal of Forecasting*, *26*(3), 448–459.

Gandar, J. M., Dare, W. H., Brown, C. R., & Zuber, R. A. (1998). Informed traders and price variations in the betting market for professional basketball games. *Journal of Finance*, *50*, 385–401.

Gandar, J. M., Zuber, R. A., & Dare, W. H. (2000). The search for informed traders in the totals betting market for National Basketball Association games. *Journal of Sports Economics*, *1*, 177–186.

Glickman, M. E., & Stern, H. S. (1998). A state-space model for National Football League scores. *Journal of the American Statistical Association*, *93*, 25–35.

Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, *26*(3), 660–670.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, *3*(3), 1–22.

Kvam, P., & Sokol, J. S. (2006). A logistic regression/Markov chain model for college basketball rankings. *Naval Research Logistics*, *53*, 788–803.

Oliver, D. (2004). *Basketball on paper: rules and tools for performance analysis*. Potomac Books.

Shirley, K. (2007). *A Markov model for basketball*. Poster presentation at New England Symposium for Statistics in Sports, Boston, MA, September 2007. http://www.amstat.org/Chapters/boston/nessis07/presentation_material/Kenny_Shirley.pdf. Accessed on 15.07.2010.

Smith, M. A., Patton, D., & Williams, L. W. (2010). Do bookmakers possess superior skills to bettors in predicting outcomes? *Journal of Economic Behavior and Organization*, *71*(2), 539–549.

Stekler, H. O., Sendor, D., & Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting*, *26*(3), 606–621.

Stern, H. (1994). A Brownian motion model for the progress of sports scores. *Journal of the American Statistical Association*, *89*, 1128–1134.

Zak, T. A., Huang, C. J., & Sigfried, J. J. (1979). Production efficiency: the case of professional basketball. *The Journal of Business*, *52*(3), 379–392.

Erik Štrumbelj is a Ph.D. student and junior researcher at the University of Ljubljana, Faculty of Computer and Information Science. His research interests include machine learning, data mining, and forecasting.

Petar Vračar received his Masters degree in computer science from the University of Ljubljana, Faculty of Computer and Information Science, where he is currently employed as a teaching assistant. His research interests include machine learning, image analysis, and sports forecasting.