Modeling and forecasting the outcomes of NBA basketball games

Hans Manner*[1]

[1]*Institute of Econometrics and Statistics, University of Cologne*

November 26, 2015

Abstract

This paper treats the problem of modeling and forecasting the outcomes of NBA basketball games. First, it is shown how the benchmark model in the literature can be extended to allow for heteroscedasticity and estimation and testing in this framework is treated. Second, time-variation is introduced into the model by introducing a dynamic state space model for team strengths. The in-sample results based on eight seasons of NBA data provide weak evidence for heteroscedasticity, which can lead to notable differences in estimated win probabilities. However, persistent time variation is only found when combining the data of several seasons, but not when looking at individual seasons. The models are used for forecasting a large number of regular season and playoff games and the common finding in the literature that it is difficult to outperform the betting market is confirmed. Nevertheless, a forecast combination of model based forecasts with betting odds can lead to some slight improvements.

Keywords: Sports forecasting, paired comparisons, NBA basketball data, heteroscedasticity, time-variation

*manner@statistik.uni-koeln.de

1 Introduction

The statistical modeling of sports data has become a large topic of research over the past decades. Detailed data of high quality have become easily available due to their publication and distribution via the internet, which allows researchers to address a variety of questions. One problem of particular interest is the prediction of the outcomes, both in terms of the final score and the winning team; see Steckler et al. (2010) for an overview. This is closely related to the issue of modeling the strength of each player or team involved in the competition of interest. The best known example of such an approach is the Elo rating in chess (Elo 1978), but similar statistical methods have been applied in many different sports. Such a strength, or rating, can be obtained by variations on the statistical method of paired comparison models by Bradley and Terry (1952) and David (1959). A notable methodological innovation was the introduction of dynamic models of paired comparison in Glickman (1993) and Fahrmeir and Tutz (1994). This approach has been applied to soccer (Fahrmeir and Tutz 1994 or Koopman and Lit 2015), chess and tennis (Glickman 1999), football (Glickman 2001, Glickman and Stern 1998), and basketball (Knorr-Held 2000), finding evidence of time-varying team/player ratings. More recently, Cattelan et al. (2013) proposed a dynamic paired comparison model based on the exponentially weighted moving average to model time-varying basketball and soccer results, whereas Baker and McHale (2015) propose a deterministic time-varying strength model to determine which English football team has been the strongest in a historical context. Percy (2015) gives an overview of stochastic processes that can be used for modeling sports data and suggests a method for dynamic updating of the model parameters.

The present paper treats the modeling and prediction of national basketball association (NBA) basketball games. The NBA is the most important and strongest professional basketball league in the world, consisting of 30 teams/franchises. With revenues of 4.6 billion US$ and an average team worth of 634 million US$ the league has a high economic relevance.

Statistical models for various aspect of basketball have been suggested in the literature. Early contributions introducing the regression based approach to basketball modeling are Stefani (1977a) and Stefani (1977b). The National Collegiate Athletic Association (NCAA) basketball tournament has been analyzed and modeled in several studies, e.g., Schwertman et al. (1991), Carlin (1996) or Harville (2003), with a focus on computing win probabilities and accurate team rankings. Stern (1994) proposes a model relying on Brownian motion that can be used to predict the outcome of a game conditional on a given score and remaining game time. A further topic that is often addressed in the literature is the home court advantage, studied in Harville and Smith (1994), Jones (2007, 2008), or Entine and Small (2008). Other studies focus more on the relevance of game statistics, such as Kubatko et al. (2007) who introduce various advanced statistics computed from box score data. Several studies, e.g., Teramoto and Cross (2010), Baghal (2012) or Page et al. (2007), explain the game outcomes using box scores and advanced statistics, in particular the *four factors*[1]. However, as this information in only known ex post, it is unclear whether these results can be exploited for forecasting purposes. A notable exception is the Markov model in Štrumbelj and Vračar (2012), in which the transition probabilities in a Markov chain model for basketball games are explained by the four factors. An interesting approach using detailed in-game data is the graphical model for match simulation by Oh et al. (2015).

The prediction of basketball games is the topic of Boulier and Stekler (1999), Caudill (2003), Loeffelhold et al. (2009), Rosenfeld et al. (2010), Stekler and Klein (2012), Štrumbelj and Vračar (2012), or Štrumbelj (2014). These predictions are done in very different settings and with quite distinct methodologies. In particular, forecasts are often based on team rankings, betting odds or statistical models. A common finding of many studies is that predictions based on betting markets are difficult to beat, thus implying efficiency of the betting markets; see also Steckler et al. (2010) and references therein on this issue.

This paper contributes to the aforementioned literature in several ways. Building on the benchmark linear model for team strengths, including parameters for the effect of the home court advantage and of playing back-to-back games, team specific volatility is introduced into the framework. The estimation and testing for heteroscedasticity is discussed. A second contribution is to consider a model for time-varying team strengths, similar to the dynamic models discussed above, in which the team strengths follow a Gaussian autoregressive process. The empirical analysis relies on a large dataset of eight NBA seasons. Estimates of teams strength and rankings, as well as the effect of the home court advantage and back-to-back games are compared across different models. Tests for heteroscedasticity are applied to the data providing some weak evidence against the assumption of equal error variances across teams. Applying the time-varying model we find only little evidence for persistent time-varying strength parameters within a single season, although the strength is persistent when pooling the data of all seasons. This

[1]The four factors are effective field goal percentage, turnovers per possession, offensive rebounding percentage, and free throw rate; see Kubatko et al. (2007) for details

is in line with the usual believe that the "hot hand" does not exist for teams; see the discussion in Camerer (1989) and Brown and Sauer (1993) on this issue. Finally, the forecasting performance of the proposed models is compared for a large number of regular season and playoff games. The model forecasts are compared to point spreads from the betting market and it turns out that this is a benchmark that is difficult to beat. The model based forecasts are also combined with the point spreads and the resulting forecast combinations show some promising results.

The rest of the paper is structured as follow. In Section 2 the methodology is explained, Section 3 presents the empirical application and some conclusions are given in Section 4. In the appendix estimation details for the dynamic state space model and additional estimation results are given. Additional and detailed empirical results for the individual seasons from 2006-2014 can be found in the online appendix of the paper.

2 Methodology

Let y_{ijk} be the difference in scores of the home team i and the away team j, where $k = 1, \ldots, n$ is the index of game k and n is the total number of games. The total number of teams is denoted by t and each team plays a total of K games, so that $n = \frac{t \times K}{2}$. A simple model for the outcome of the game is

$$y_{ijk} = \lambda + \alpha(B_i - B_j) + \beta_i - \beta_j + e_{ijk}, \tag{1}$$

where λ denotes the (constant) home advantage, B_i is a dummy variable indicating whether team i plays back-to-back games, i.e., games on two consecutive days, with α the corresponding effect, and β_i and β_j denote the strength of teams i and j, respectively.[2] The error term e_{ijk} is assumed to be normally distributed with mean 0 and variance σ^2. Harville (2003) suggests accounting for the discreteness of the observed scores. However, normality tests and the results in Stern (1994) suggest that the residuals from model (1) and its extensions below are normally distributed. Furthermore, normality of the error terms implies that the correction for blowout victories proposed in Harville (2003) is not necessary and would, in fact, lead to inefficient estimates given the fact that under normality ordinary least squares (OLS) is equivalent to the (asymptotically efficient) maximum likelihood estimator. We can state the model in matrix form letting $\boldsymbol{y}$ be the $n \times 1$ vector

[2] Here we made the assumption that the effect of playing back-to-back games is the same for the home and away team.

of spreads, $\boldsymbol{e}$ the $n \times 1$ vector of errors, $\boldsymbol{\beta} = [\lambda\ \alpha\ \beta_1 \dots \beta_t]'$ the vector of coefficients and $\boldsymbol{X}$ the $n \times (t+1)$ design matrix. A typical row of this matrix has 1 as its first element (for the home advantage), $B_i - B_j$ in the second column, 1 in column $i+2$ and -1 in column $j+2$ in the case that it corresponds to a game of team i (home) against team j (away). The remaining elements are equal to 0. Then the model is compactly given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}. \tag{2}$$

However, the matrix $\boldsymbol{X}$ is not of full rank, so for estimation one can remove the third column. This corresponds to the normalizing restriction $\beta_1 = 0$, meaning that the strength of the first team is set equal to zero. Without this restriction the parameter vector $\boldsymbol{\beta}$ cannot be identified, as adding a constant to each team strength leads to an equivalent model. The parameters can be then estimated by OLS:

$$\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{3}$$

2.1 Heteroscedasticity

The model above assumes constant variance of the error term, i.e., $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, where $\boldsymbol{I}$ is the $n \times n$ identity matrix. Here we relax this assumption. Let the strength of team i in game k be given by

$$S_{ik} = \beta_i + e_{ik}, \tag{4}$$

where β_i is the constant component of the team strength and $e_{ik} \overset{iid}{\sim} N(0, \sigma_i^2)$ the team specific error term. Thus the strength of a team in a specific game consists of a constant component and an error term. A larger value of the error variance σ_i^2 implies that the corresponding team shows a more volatile performance. Then the outcome of the game is modeled as

$$y_{ijk} = \lambda + \alpha(B_i - B_j) + S_{ik} - S_{jk} = \lambda + \alpha(B_i - B_j) + \beta_i - \beta_j + \underbrace{e_{ik} - e_{jk}}_{e_{ijk}}. \tag{5}$$

Consequently, the baseline model (1) is obtained when $\sigma_i^2 = \sigma^2/2$ for all i. In matrix notation the model is the same as (2), but with $\mathrm{Cov}(\boldsymbol{e}) = \Omega \neq \sigma^2 \boldsymbol{I}$. The matrix Ω is diagonal with typical element $\sigma_i^2 + \sigma_j^2$, corresponding to a game between teams i and j.

The model can be estimated in two ways: Maximum likelihood estimation (MLE) or feasible generalized least squares (FGLS); see Greene (2011) for details on GLS estimation. MLE is straightforward since $e_{ijk} \sim N(0, \sigma_i^2 + \sigma_j^2)$ and the errors are independent. To estimate the model by FGLS first estimate (2) by OLS to obtain the residual vector $\hat{\boldsymbol{e}}$. Next, run the regression

$$\hat{\boldsymbol{e}}^2 = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{6}$$

where $\hat{\boldsymbol{e}}^2$ is the vector of squared residuals and the $n \times t$ matrix $\boldsymbol{Z}$ has a typical row with entries of 1 in columns i and j if the observation corresponds to a game between teams i and j and zeros in the remaining columns. The estimated parameter vector $\hat{\boldsymbol{\gamma}}$ in fact gives estimates for the team specific variances σ_i^2. The fitted values from (6), say $\hat{\sigma}_{ijk}^2$, make up the elements on the main diagonal of our estimate for the covariance matrix of the error terms $\hat{\Omega}$. Then the FGLS estimator is given by

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\boldsymbol{X}'\hat{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\Omega}^{-1}\boldsymbol{y}. \tag{7}$$

A natural question is whether the model should be estimated by MLE or by FGLS, which differ in finite samples. FGLS has the advantage that it is easy to compute and does not require numerical optimization, whereas MLE is appealing due to the asymptotic optimality properties of maximum likelihood estimators when the model is correctly specified. However, no statement can be made which estimator is preferable in finite samples.

Consider testing the null hypothesis of homoscedasticity, i.e., a constant error variance across teams,

$$H_0 : \sigma_i^2 = \sigma_j^2 \text{ for all } i \neq j. \tag{8}$$

There are two ways we can test this hypothesis. First, one could estimate model (5) by MLE and additionally estimate the model under the restriction of homoscedasticity. Let LL_0 be the log-likelihood under H_0 and LL_1 under the alternative. Then we can test H_0 using

$$LR = 2(LL_1 - LL_0), \tag{9}$$

which follows a χ^2 distribution with $t-1$ degrees-of-freedom under the null. Alternatively, we can base our test on the regression (6). Let SSR_1 be the sum-of-squared residuals from this model and let SSR_0 be the residuals from regressing $\hat{\boldsymbol{e}}^2$ on a constant. Then we can test H_0 with the F-statistic

$$F = \frac{(SSR_0 - SSR_1)/(t-1)}{SSR_1/(n-t)}, \tag{10}$$

which is distributed $F(t-1, n-t)$.

In general, one may be interested in computing the probability that team i (the home team) wins a specific game. This can be computed as

$$\begin{aligned}
P(\text{Team } i \text{ wins}) &= P(y_{ijk} > 0) = P(\lambda + \alpha B_i + S_{it} > \alpha B_j + S_{jk}) \\
&= P(\lambda + \alpha B_i + \beta_i + e_{ik} > \beta_j + \alpha B_j + e_{jk}) \\
&= P(e_{jk} - e_{ik} < \lambda + \alpha(B_i - B_j) + \beta_i - \beta_j) \\
&= P\left(\frac{e_{jk} - e_{ik}}{\sqrt{\sigma_i^2 + \sigma_j^2}} < \frac{\lambda + \alpha(B_i - B_j) + \beta_i - \beta_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right) \\
&= \Phi\left(\frac{\lambda + \alpha(B_i - B_j) + \beta_i - \beta_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right), \qquad (11)
\end{aligned}$$

where Φ denotes the CDF of the standard normal distribution.

2.2 Dynamic Modeling

In this section we consider a model in which the strength of team i is a time-varying latent process. It is assumed to follow a Gaussian autoregressive process of order one. Let the strength parameter be indexed by $k_i = 1, \ldots, 82$[3], i.e., we now have β_{i,k_i}. The outcome of the game in this context is modeled by

$$y_{ijk} = \lambda + \alpha(B_i - B_j) + \beta_{i,k_i} - \beta_{j,k_j} + e_{ijk}, \qquad (12)$$

where $e_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$. The time-varying team strength evolves as

$$\beta_{i,k_i} = \mu_i + \phi_i \beta_{i,k_i-1} + \eta_{k_i}, \qquad (13)$$

where $\eta_{k_i} \sim N(0, \sigma_{\eta_i}^2)$. Although this is a state space model and β_{i,k_i} is unobservable the estimation is relatively straightforward due to the fact that both e_k and η_{k_i} are normally distributed. The key difference to a standard state space model in time series analysis is the fact that the observations are not equidistant in calendar time, and therefore the evolution of the strength is defined from game to game.[4] Nevertheless, the Kalman filter can be applied to estimate the model parameters and the strengths of the teams. The

[3]Note that each team plays 82 games per season, with the exception of lockout seasons such as the 2011-2012 season. When the model is applied to multiple seasons the number of games per team changes accordingly.

[4]A model in which strength evolves in calendar time was also considered in a preliminary analysis.

details on how this is done for this specific model are given in the appendix. We impose one set of restrictions to the model in order to reduce the number of free parameters, namely we restrict ϕ_i to be the same for all teams. Furthermore, we also consider imposing the restriction that $\sigma^2_{\eta_i}$ is the same for all teams, addressing the issue whether heteroscedasticity is still an issue when allowing for time-varying strength parameters. Again, a standard likelihood ratio test can be used to test this restriction.

When analyzing the data of multiple seasons we want to allow for a faster adjustment of the team strength at the beginning of the season, due to the fact that trades, retirements and draft picks are likely to results in significant changes in team strengths from one season to another. Here we follow the suggestion of Koopman and Lit (2015) and replace the distribution of η_{k_i} by

$$\eta_{k_i} \overset{iid}{\sim} N(0, \sigma^2_{\eta_i} + \sigma^2_{FG} I_{\{FG_i\}}), \tag{14}$$

where the indicator $I_{\{FG_i\}}$ is equal to 1 for the first game team i plays in each season. As noted by Koopman and Lit (2015), when the team strength has high persistence this will lead to breaks in the process.

3 Application

In this section we apply the models proposed in Section 2 to a large data set of NBA games covering the Seasons 2006-2007 until 2013-2014, thus a total of eight NBA seasons. The data was obtained from www.nbastuffer.com. Besides the outcomes of the games and betting odds[5], the data set contains further information that was not used in this study such as the box score, the starting lineups and some advanced basketball statistics.

In a typical regular season each of the 30 teams plays 82 games, resulting in a total of 1230 regular season games. An exception is the 2011-2012 lockout season in which each team played 66 games, implying a total of 990 regular season games. Furthermore, during the 2012-2013 season as a result of the bombing at the Boston marathon the game Boston vs. Indiana needed to be rescheduled and was eventually not played.

The rest of this section is structured as follows. In Section 3.1 we present and discuss the in-sample results. Section 3.2 compares the forecasting performance of the models for both regular season and playoff games.

[5]Based on www.scoresandodds.com.

3.1 In-sample results

Here we consider the modeling of the regular season data for all available seasons. The results of the static models are discussed in Section 3.1.1 and the results of the dynamic model can be found in Section 3.1.2.

3.1.1 Static models

In this section we address the questions whether the variance of the team strength differs between teams and whether the incorporation of heteroscedasticity influences the estimation of the team strength and the ranking of the teams. Furthermore, we provide estimates of the home court advantage and the effect of playing back-to-back games. An example illustrates how these factors affect the estimated winning probabilities The analysis was conducted for each individual season from 2006 to 2014 and for the pooled data including (team specific) season dummies to allow the strength of the teams to vary between seasons. We only report the results for the 2013-2014 season and for the pooled estimation. The complete results can be found in the online appendix of the paper.

Table 1 presents the estimates for the home advantage and the effect of back-to-back games, as well as the p-values of the F-test and likelihood ratio test for the null hypothesis of homoscedasticity given in equations (9) and (10). The likelihood ratio test is additionally applied for the dynamic model characterized by equations (12) and (13). The test results give some evidence in favor of heteroscedasticity, although for the pooled data the homescedasticity cannot be rejected at the 1% significance level. The results for the remaining seasons, to be found in the online appendix, are mixed. Considering the fact that we perform the tests over eight seasons, using a simple Bonferroni adjustment for each test individually suggests rejection when the p-value is below $0.05/8 = 0.00625$ when testing at $\alpha = 0.05$. This suggests rejection only in 3 out of 8 seasons. Taken jointly these results suggest that there is only weak evidence in favor of heteroscedasticity. The effect of the home advantage is estimated to be around 2.7 points per game, whereas the playing back-to-back games on average results in a disadvantage of about 1.8 points. These results are robust across the different models and seasons.

Additionally, several normality tests were applied on the estimated residuals of the different models. For each season individually normality cannot be rejected, whereas tests applied to the residuals of the pooled data provide some evidence against normality. The detailed results can be found in the online appendix.

The estimated team strengths, rankings and estimated variances for the 2013-2014

Table 1: Home advantage, effect of back-to-back games and heteroscedasticity tests

	OLS	GLS	MLE	Dynamic
		2013-2014		
Home	2.29	2.22	2.29	2.21
	(0.35)	(0.32)	(0.33)	(0.33)
B2B	-1.85	-1.70	-1.69	-1.73
	(0.66)	(0.60)	(0.61)	(0.62)
Het.	-	0.0004	0.0000	0.0000
		2006-2014		
Home	2.70	2.69	2.69	2.70
	(0.12)	(0.12)	(0.12)	(0.12)
B2B	-1.87	-1.86	-1.86	-1.88
	(0.23)	(0.22)	(0.22)	(0.22)
Het.	-	0.0278	0.0044	0.0275

Note: Table 1 presents the estimated homecourt avantage, the effect of playing back-to-back games (B2B) with standard error in parentheses and the p-values of the tests for heteroscedasticity (Het.). The results are based on the models defined in equations (1), (5) and (12), denoted by OLS, GLS/MLE and Dynamic, respectively. GLS and MLE refer to the estimation method of the heteroscedastic model (5).

season can be found in Table 4 in Appendix B. The parameter estimates show some differences between the different estimators and some slight differences in team rankings emerge when allowing for heteroscedasticty. Looking at the range of estimated team strengths it can be seen that the difference between the best (San Antonio) and the worst (Philadelphia) team in the league implies an expected point difference of about 18 points. Looking at the variance estimates themselves no clear pattern emerges. High variances are possible both for successful and unsuccessful teams.

In order to get a feeling of the implications for predicting the outcomes of games based on the different models we computed the winning probabilities for a few hypothetical games in the 2013-2014 season. The teams we consider are ones characterized by high estimated variances, (New York, Chicago, and Philadelphia) and teams with low team estimated variances (Orlando, Milwaukee, Dallas). Their estimated strengths and variances can be found in Table 4. Note that the estimated error variance for the homoscedastic model is $\hat{\sigma}^2_{OLS} = 136.6$. In Table 2 we report the estimated win probabilities of Team 1 vs. Team 2 in a number of settings, computed using equation (11). In particular,

Table 2: Predicted winning probabilities

Team 1	Team 2	$P_{Hom,1@2}$	$P_{Het,1@2}$	$P_{Hom,2@1}$	$P_{Het,2@1}$
New York	Chicago	0.3473	0.3941	0.5006	0.5042
B2B		0.4079	0.4343	0.564	0.5452
	B2B	0.2903	0.3551	0.4371	0.4631
Orlando	Milwaukee	0.5079	0.5013	0.6606	0.7651
B2B		0.5712	0.6061	0.7169	0.8386
	B2B	0.4444	0.3965	0.6004	0.6762
New York	Philadelphia	0.7317	0.6766	0.8443	0.7711
B2B		0.7816	0.7134	0.8794	0.8016
	B2B	0.6766	0.6381	0.803	0.7381
Dallas	Milwaukee	0.7786	0.9354	0.8773	0.989
B2B		0.8231	0.9643	0.9068	0.995
	B2B	0.7283	0.8909	0.8418	0.9775

Note: Table 2 presents the predicted probability that Team 1 will beat Team 2 with either team playing at home. 'Hom' stands for the homoscedastic model and 'Het' for the heteroscedastic model. B2B indicates that the respective team is assumed to play back-to-back games. The numbers are based on the estimates for the 2013-2014 season that can be found in Table 4.

we compare the probabilities for the homoscedastic baseline models estimated by OLS and the heteroscedastic model estimated by MLE. We consider the situation that either team plays at home and additionally that each team plays back-to-back (B2B) games. The probabilities based on the homescedastic and heteroscedastic models are very similar when the win probabilities are close to 0.5, but the probabilities differ significantly, up to 0.17, when one team is more likely to win. The home court advantage can lead to differences in estimated win probabilities of up to 0.26, whereas the effect back-to-back games can change the win probability up to about 0.1. While these examples consider a rather extreme situation when both teams have either high or low variances, it shows that heteroscedasticity can affect estimated win probabilities quite strongly. Furthermore, these numbers give an impression of the importance of playing home/away and back-to-back games not in terms of expected difference in the spread, but in terms of winning probabilities.

3.1.2 Dynamic modeling

In order to shed some light on the question of momentum in team strength we treat the residuals of the static model as panel data for each team over the course of the individual seasons and perform the Lagrange-multiplier test for autocorrelation by Baltagi and Li (1998). In all cases the null hypothesis of no-autocorrelation cannot be rejected[6]. This provides some initial evidence against persistent and predictable time-variation in team strengths.

The next step in the analysis is the estimation of the dynamic state space model from Section 2.2. Intuitively this model seems a reasonable approach, as one would expect the strengths of teams to change throughout the course of a season due to injuries, trades, changes in coaching and team chemistry, etc. Considering only the individual seasons, however, the log-likelihood of the dynamic and static models are basically identical for all seasons and the point estimates for the persistence parameter ϕ is always close to 0. Furthermore, for individual seasons the smoothed and filtered estimates of the path of the team strengths look rather erratic and do not suggest any persistence.

One explanation of these results may be that the persistence parameter ϕ may be difficult to estimate with the limited number of games in each season. Therefore, the data from all seasons was pooled and the extended model allowing for an increased error variance at the beginning of each season as described in equation (14) was used. For this model (assuming homescedasticity across teams) the parameter estimates were $(\hat{\phi}, \hat{\sigma}^2, \hat{\sigma}^2_\eta, \hat{\sigma}^2_{FG}) = (0.9942, 129.8601, 0.0966, 11.3845)$, with estimated standard errors 0.0012, 1.9870, 0.0326, and 2.1376, respectively. The autoregressive parameter is now close to 1 indicating a strong degree of persistence. This can mainly be explained by a large degree of persistence within each season. Figures 1 and 2 in the appendix show the smoothed estimates of the time-varying team strengths together with the static results. The static strengths β_i have been normalized to add up to zero as suggested by Knorr-Held (2000) using the formula $\tilde{\beta}_i = \hat{\beta}_i - \frac{1}{30}\sum_{j=1}^{30}\hat{\beta}_j$. Similarly, the constants in the dynamic model α_i have been normalized to add up to zero, which makes the dynamic and static strengths comparable. Several things can concluded from the graphs. The change in the strengths at the beginning of each season is notable and confirms that the increase in the variance for the first game of each season basically introduces the possibility of a structural break in strengths. Furthermore, the static and dynamic strengths are very close to each other in most cases and lead to very similar rankings of the teams. In fact, for many teams

[6]Detailed results for all unreported findings in this section are available from the author upon request.

and in many seasons the time-variations is not very pronounced as the variation of the strength within a single season is very small compared to the between-team variation. There are several exceptions to this. For example, in the 2008-2009 season the strength of the Boston Celtics declines steadily, which can be explained by a sensational start of the season, being 27-2, and a normalization of the performance thereafter. Another notable example is the steady improvement of the young Oklahoma City Thunder in the 2008-2009 season. Finally, there are several instances when well performing teams such as the San Antonio Spurs become weaker towards the end of the regular season. This typically can be explained by the fact that these teams are often qualified for the playoff early in the season and decide to give their key players more rest before the playoffs begin.

3.2 Predictability

In this section we consider the problem of forecasting the game outcomes using the models described above. This is done for regular season and for playoff games. The forecasts are evaluated using three criteria. The first criterion is the mean square prediction error (MSE):

$$MSE = \sum_{k=1}^{n^*} (y_{ijk} - \hat{y}_{ijk})^2,$$

where n^* is the number of out-of-sample observations. The second criterion is the mean absolute prediction error (MAE),

$$MAE = \sum_{k=1}^{n^*} |y_{ijk} - \hat{y}_{ijk}|,$$

and the third criterion is the fraction of games in which the correct winner was predicted. Whereas the MSE is the obvious choice for the loss function given the fact that the error terms can safely be considered to be Gaussian, the other two criteria are easy to interpret.

The models considered in the forecasting exercise are the homoscedastic baseline model (OLS), the heteroscedastic model (Het.) estimated by MLE and the dynamic state space model (Dyn.). As a benchmark the Las Vegas opening spreads (Spr.) for bets on the games are considered. Furthermore, for all models we consider the combined forecasts of the model based forecasts with the betting spreads. The forecasts are combined with equal weights, as a preliminary analysis suggested that the two types of forecasts have approximately the same variances and are highly correlated (> 0.9). Therefore more

Table 3: Forecast evaluation

	OLS	Het.	Dyn.	Spr.	OLS-Spr.	Het.-Spr.	Dyn.-Spr.
Regular Season							
MSE	142.21†	142.25†	143.13†	**137.49**	137.93	137.93	138.37
MAE	9.36†	9.36†	9.38†	**9.17**	9.20	9.20	9.22
Correct	0.682	0.681	0.684	0.692	0.693	0.692	**0.694**
Playoffs							
	OLS	Het.	Dyn.	Spr.	OLS-Spr.	Het.-Spr.	Dyn.-Spr.
MSE	151.24†	151.14†	150.85	148.96	148.046	148.06	**147.69**
MAE	9.69†	9.68†	9.68	**9.50**	9.54	9.53	9.53
Correct	0.687	0.684	0.686	0.680	0.686	**0.689**	0.683

Note: Table 3 gives the predictive mean-square-error (MSE), mean-absolute-error (MAE) and fraction of correctly predicted outcomes combined for all seasons from 2006 to 2014. The regular season results are based on the second half of each season, with recursively estimated model parameters using all previous games of each season. The playoff results are based on parameter estimates using all regular season games each year. OLS refers to the homoscedastic model in (1), Het. to the heteroscedastic model in (5), Dyn. to the dynamic state space model in (12), and Spr. to the Las Vegas opening spreads. The remaining three columns refer to equally weighted forecast combinations. The results for the best performing model are presented in bold. A † implies that the corresponding model is not included in the 95% model confidence set.

sophisticated weighting schemes do not appear to be sensible here; see Timmermann (2006) for extensions.

In order to decide whether the differences in forecasting accuracy across models are statistically significant, the model confidence set (MCS) by Hansen et al. (2011) is computed based on the MSE and MAE loss functions. The MCS is a set of models whose forecasting performance is not significantly different considering a certain loss function and it can be seen as an analogue to a confidence interval for competing (non-nested) models. Thus it acknowledges the fact that it is unlikely that a single model outperforms all the others, but that there are multiple models that perform equally well. The MCS is determined using a sequence of hypothesis tests. It eliminates inferior models based on the criterion of interest. P-values for the sequential tests are determined by bootstrap procedure as described in Hansen et al. (2011) and references therein. A size of 5% and 10,000 bootstrap samples are used to compute the MCS.

The forecasting performance for the regular season data is analyzed as follows. The first half of the regular season data, 615 games in a typical season, are used as the in-sample period, whereas the remaining games constitute the out-of-sample period. The models are re-estimated using an expanding window scheme to produce forecasts for the full out-of-sample period. This is done for each season separately, but due to the presence of season-dummies the results are identical to using multi-season data for the static models. For the dynamic model the use of the multi-season data lead to significantly worse results in terms of forecasting performance[7]. For the forecast evaluation of the playoff games the complete regular season data is used as the training period, but the models are not re-estimated during playoff period. In the case of the dynamic state space, however, the information set is updated throughout the playoffs and the predicted values based on the Kalman filter are used as forecasts.

The results combining the predictions for all eight season are presented in Table 3[8], with the results for the best performing model in each case bold. A † indicates that the respective model is excluded from the model confidence set, indicating that its loss is significantly worse than that of the best performing model. For the regular season games the betting odds provide the best predictions in terms of MSE and MAE, although the combined forecasts are very close and give very slight improvements for predicting the correct outcomes. The pure model based forecasts perform significantly worse than the betting odds. A similar picture emerges for the playoff games. Again, the Las Vegas spreads provide forecasts that are hard to beat, but combining these forecasts with the model based approaches can result in small improvements of the forecasts.

Overall, about 69% of the game outcomes can be predicted correctly and it seems questionable that much better forecasts are possible, as a certain amount of randomness/unpredictability is an inherent part of sports. Furthermore, the Las Vegas spread remains a benchmark for prediction that appears to be very difficult to beat. Comparing the mean square prediction errors with the in-sample residuals variance, which is estimated at about 130 for the different model specifications, it is clear that there is very little room for improvement unless powerful predictors for game outcomes can be found.

4 Conclusion

In this paper we have reconsidered the modeling of team strength in professional basketball. The standard model was extended by allowing for team specific error variances

[7]Results are available upon request.

[8]The results for the individual season can be found in the online appendix.

and time-variation in team strength. These models were applied to the NBA games in all eight seasons in the period 2006 until 2014. The results of the in-sample estimation suggest some evidence of heteroscedasticity. Furthermore, the evidence for persistent time-variation in teams strengths is much weaker than one would expect given injuries, trades, and other factor influencing the team composition and chemistry. This is confirmed by the non-rejection of a test for no autocorrelation on the residuals of the static models.

Besides the methods presented in this paper several other models were considered that were not able to improve the model fit. In particular, a model treating offensive and defensive strength separately in both a static and dynamic setting did not yield improvements in fit. A random walk model was considered to avoid the estimation of the persistence parameter. However, the variance of the errors of the state-equation was estimated at zero, implying a constant strength. Furthermore, instead of the dynamic state space model, an autoregressive observation driven approach for team strength in which the residuals of the previous game were allowed to drive the current team strength was considered. Given the lack of evidence for time-variation within single seasons, it is not surprising that such a model could not outperform simpler static models.

The forecasting performance of the models was evaluated using regular season and playoff games over all eight seasons. These finding confirm the common theme in the literature on sports forecasting: it is difficult to beat the betting markets, which indicates that they are efficient. However, combining the model based forecasts with betting spreads can results in some improvements and the model confidence sets imply that the combined forecasts are statically not worse than the one based solely on betting spreads.

Future research should address the question whether advanced basketball statistics suggested in Kubatko et al. (2007) can be used to improve model based forecasts and whether these statistics themselves are predictable. Furthermore, more detailed information concerning injuries or suspensions of key players can be incorporated into the models for forecasting purposes. Finally, it could be interesting to search for factors that can explain the different team variances.

A Implementation of the Kalman filter

The latent state vector of interest is β_{i,k_i} for $i = 1, \ldots, 30$, each of length equal to the number of games played by each team. For analyzing a single season, e.g., $k_i = 1, \ldots, 82$. The set of hyperparameters to be estimated consists of μ_i for $i = 1, \ldots, 30$, ϕ, σ^2 and $\sigma^2_{\eta_i}$ for $i = 1, \ldots, 30$. In case of homoscedasticity the latter parameter is the same for each team. Finally, when allowing for a break in strengths at the beginning of each season one additionally has to estimate σ^2_{FG} and the steps below have to be adjusted accordingly. Let $\beta_{i,k_i|k_i-1}$ be the predicted team strength of team i for game k_i conditional on the information at game $k_i - 1$, whereas $\beta_{i,k_i|k_i}$ denotes the updated strength conditional on information up to game k_i. The variance of β_{i,k_i} conditional on information at game $k_i - 1$ is denoted as $P_{i,k_i|k_i-1}$, whereas the updated variance of team i is $P_{i,k_i|k_i}$. Then the steps of the Kalman filter for game k between teams i and j with outcome y_{ijk}, being games k_i and k_j for the teams, respectively, are as follows.

Prediction step:

$$
\begin{aligned}
\beta_{i,k_i|k_i-1} &= \mu_i + \phi\beta_{i,k_i-1|k_i-1} \\
\beta_{i,k_j|k_j-1} &= \mu_j + \phi\beta_{i,k_j-1|k_j-1} \\
P_{i,k_i|k_i-1} &= \phi^2 P_{i,k_i-1|k_i-1} + \sigma^2_{\eta_i} \\
P_{i,k_j|k_j-1} &= \phi^2 P_{i,k_j-1|k_j-1} + \sigma^2_{\eta_j}
\end{aligned}
$$

Observation step:

$$
\begin{aligned}
\hat{y}_{ijk} &= \lambda + \alpha(B_i - B_j) + \beta_{i,k_i|k_i-1} - \beta_{i,k_j|k_j-1} \\
V_{ijk} &= P_{i,k_i|k_i-1} + P_{i,k_j|k_j-1} + \sigma^2 \\
\hat{e}_{ijk} &= y_{ijk} - \hat{y}_{ijk}
\end{aligned}
$$

Updating step:

$$
\begin{aligned}
\beta_{i,k_i|k_i} &= \beta_{i,k_i|k_i-1} + \hat{e}_{ijk} P_{i,k_i|k_i-1}/V_{ijk} \\
\beta_{i,k_j|k_j} &= \beta_{i,k_j|k_j-1} - \hat{e}_{ijk} P_{i,k_j|k_j-1}/V_{ijk} \\
P_{i,k_i|k_i} &= P_{i,k_i|k_i-1} - P^2_{i,k_i|k_i-1}/V_{ijk} \\
P_{i,k_j|k_j} &= P_{i,k_j|k_j-1} - P^2_{i,k_j|k_j-1}/V_{ijk}
\end{aligned}
$$

The initial values are set to $\beta_{i,1|0} = \mu_i/(1-\phi)$ and $P_{i,1|0} = \sigma^2_{\eta_i}/(1-\phi^2)$. The log-likelihood contribution of the kth game is given by

$$\ln L_k = \frac{1}{2}\ln(2\pi) + \frac{1}{2}\ln(V_{ijk}) + \frac{\hat{e}^2_{ijk}}{2V_{ijk}}.$$

This likelihood has to be maximized numerically over the set of hyperparameters to obtain the maximum likelihood estimator of the model. Standard errors can be obtained straightforwardly by numerical estimates of the information matrix.

Finally, if one is interested in the estimates of the strength conditional on the information of the whole sample the Kalman smoother should be applied. Smoothed state estimates, denoted as $\beta_{i,k_i|K}$, are obtained by iterating the following recursion on the whole sample going from the last to the first game:

$$\beta_{i,k_i|K} = \beta_{i,k_i|k_i} + \phi\frac{P_{i,k_i|k_i}}{P_{i,k_i+1|k_i}}(\beta_{i,k_i+1|K} - \beta_{i,k_i+1|k_i}).$$

B Teams strengths and rankings

Table 4: Ranking, strength and team specific variances 2013-2014

2013-2014	rank OLS	rank FGLS	rank MLE	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FGLS}$	$\hat{\beta}_{MLE}$	$\hat{\sigma}^2_{GLS}$	$\hat{\sigma}^2_{MLE}$
San Antonio	1	1	1	8.84	9.64	9.66	79.24	70.32
LA Clippers	2	2	2	7.92	8.21	8.27	82.13	74.79
Oklahoma City	3	3	3	7.48	7.16	7.36	59.27	69.95
Houston	4	5	5	5.86	6.16	6.37	70.48	54.04
Golden State	5	4	4	5.82	6.39	6.50	56.22	55.69
Portland	6	7	7	5.20	5.14	5.25	49.87	50.38
Miami	7	6	6	5.11	6.10	6.20	64.17	68.75
Indiana	8	9	9	4.53	4.65	4.78	88.06	85.47
Phoenix	9	8	8	3.95	4.71	5.00	43.55	36.20
Minnesota	10	12	11	3.94	4.17	4.16	75.88	63.61
Dallas	11	10	10	3.68	4.34	4.27	22.79	19.12
Toronto	12	11	12	3.28	4.26	4.14	0.77	2.45
Memphis	13	13	13	2.93	2.52	2.84	43.24	65.23
Chicago	14	14	14	1.81	2.07	2.21	95.17	111.65
Washington	15	15	15	1.47	1.60	1.78	39.32	50.19
Charlotte	16	16	16	0.04	0.82	0.84	67.37	84.83
Atlanta	17	19	20	0.00	0.00	0.00	46.42	59.23
New York	18	20	19	-0.46	-0.05	0.10	154.23	157.26
Denver	19	17	17	-0.58	0.30	0.45	98.23	98.42
Brooklyn	20	18	18	-0.63	0.17	0.22	109.36	89.03
Sacramento	21	21	21	-1.07	-0.33	-0.10	76.78	73.75
New Orleans	22	22	22	-1.20	-1.11	-1.30	23.80	24.21
Cleveland	23	23	23	-2.94	-2.54	-2.43	85.04	74.78
Detroit	24	24	24	-3.28	-3.17	-3.05	71.65	64.93
Boston	25	25	25	-4.05	-3.25	-3.09	46.00	36.98
LA Lakers	26	26	26	-4.31	-4.08	-3.99	102.43	101.73
Orlando	27	27	27	-4.99	-4.65	-4.67	32.04	24.60
Utah	28	28	28	-5.36	-5.19	-5.05	77.77	85.48
Milwaukee	29	29	29	-7.51	-7.02	-6.98	10.07	15.82
Philadelphia	30	30	30	-9.91	-9.72	-9.57	96.27	101.89

Note: Table 4 presents the estimated ranking, team strengths and team specific error variances based on models (1) and (5) in the paper. The heteroscedastic model is estimated either by FGLS or by MLE.

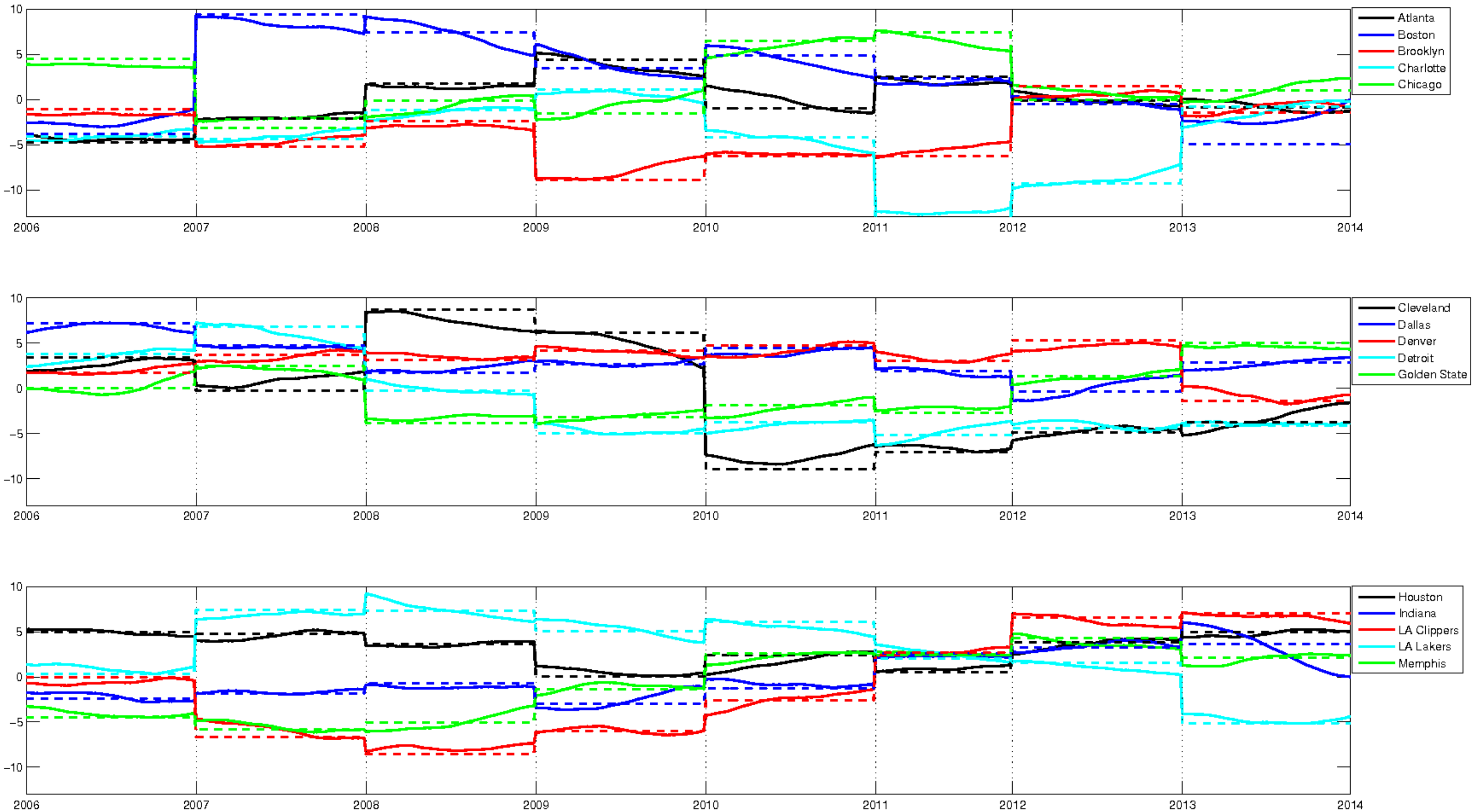

Figure 1: Team strengths over time 1 (Dynamic model solid lines, static model dashed lines)

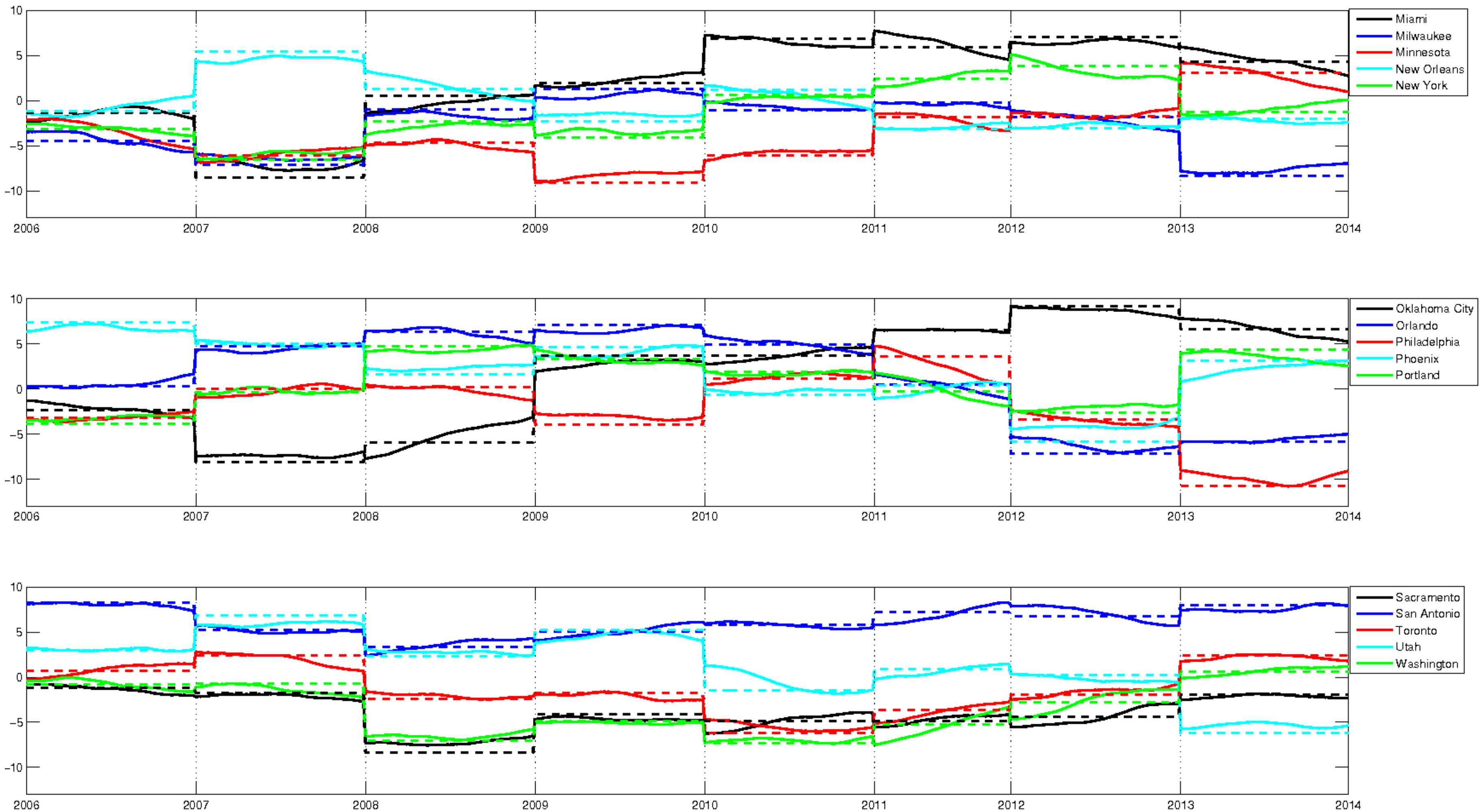

Figure 2: Team strengths over time 2 (Dynamic model solid lines, static model dashed lines)

References

Baghal, T. (2012). Are the "four factors" indicators of one factor? An application of structural equation modeling methodology to NBA data in prediction of winning percentage. *Journal of Quantitative Analysis in Sports 8*(1), Article: 4.

Baker, R. D. and I. G. McHale (2015). Time varying ratings in association football: the all-time greatest team is... *Journal of the Royal Statistical Society: Series A 178*(2), 481–492.

Baltagi, B. H. and Q. Li (1998). Testing AR(1) against MA(1) disturbances in an error component model. *Econometrica 66*(1), 47–78.

Boulier, B. L. and H. O. Stekler (1999). Are sports seedings good predictors?: An evaluation. *International Journal of Forecasting 15*, 83–91.

Bradley, R. A. and M. E. Terry (1952). The rank analysis of incomplete designs, 1. The method of paired comparisons. *Biometrika 39*, 324–345.

Brown, W. O. and R. D. Sauer (1993). Does the basketball market believe in the hot hand? Comment. *American Economic Review 83*, 1377–1386.

Camerer, C. F. (1989). Does the basketball market believe in the hot hand? *American Economic Review 79*, 1257–1261.

Carlin, B. P. (1996). Improved NCAA basketball tournament modeling via point spread and team strength information. *The American Statistician 50*, 39–43.

Cattelan, M., C. Varin, and D. Firth (2013). Dynamic Bradley–Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C 62*(1), 135–150.

Caudill, S. B. (2003). Predicting discrete outcomes with the maximum score estimator: The case of the NCAA men's basketball tournament. *International Journal of Forecasting 19*, 313–317.

David, H. A. (1959). Tournaments and paired comparisons. *Biometrika 46*, 139–149.

Elo, A. E. (1978). *The rating of chess players past and present.* New York: Arco.

Entine, O. A. and D. S. Small (2008). The role of rest in the NBA home-court advantage. *Journal of Quantitative Analysis in Sports 4*(2), Article: 6.

Fahrmeir, L. and G. Tutz (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association 89*, 1438–1449.

Glickman, M. E. (1993). *Paired comparison models with time-varying parameters*. Phd dissertation, Department of Statistics, Harvard University, Cambridge.

Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. *Applied Statistics 48*, 377–394.

Glickman, M. E. (2001). Dynymic paired comparison models with stochastic variances. *Journal of Applied Statistics 28*, 673–689.

Glickman, M. E. and H. S. Stern (1998). A state-space model for national football league scores. *Journal of the American Statistical Association 93*, 25–35.

Greene, W. (2011). *Econometric Analysis*. Pearson Education.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*, 453–497.

Harville, D. A. (2003). The selection of seeding of college basketball or football teams for postseason competition. *Journal of the American Statistical Association 98*, 17–27.

Harville, D. A. and M. H. Smith (1994). The home-court advantage: How large is it and does it vary from team to team. *The American Statistician 48*, 22–29.

Jones, M. B. (2007). Home advantage in the NBA as a game-long process. *Journal of Quantitative Analysis in Sports 3*(4), Article: 2.

Jones, M. B. (2008). A note on team-specific home advantage in the NBA. *Journal of Quantitative Analysis in Sports 4*(3), Article: 5.

Knorr-Held, L. (2000). Dynamic ratings of sports teams. *The Statistician 49*, 261–276.

Koopman, S. J. and R. Lit (2015). A dynamic bivariate poisson model for analyzing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society A 178*, 167–186.

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

Loeffelhold, B., E. Bednar, and K. W. Bauer (2009). Predicting NBA games using neural networks. *Journal of Quantitative Analysis in Sports 5*(1), Article: 7.

Oh, M.-H., S. Keshri, and G. Iyengar (2015). Graphical model for basketball match simulation. In *MIT Sloan Sports Analytics Conference*.

Page, G. L., G. W. Fellingham, and C. S. Reese (2007). Using box-scores to determine a positions's contribution to winning basketball games. *Journal of Quantitative Analysis in Sports 3*(4), Article: 1.

Percy, D. F. (2015). Strategy selection and outcome prediction in sport using dynamic learning for stochastic processes. *Journal of the Operational Research Society 66*, 1840–1849.

Rosenfeld, J. W., J. I. Fisher, D. Adler, and C. Morris (2010). Predicting overtime with the pythagorean formula. *Journal of Quantitative Analysis in Sports 6*(2), Article: 1.

Schwertman, N. C., T. A. McCready, and L. Howard (1991). Probability models for the NCAA regional basketball tournaments. *The American Statistician 45*, 35–38.

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

Stefani, R. T. (1977a). Football and basketball prediction using least squares. *IEEE Transactions on Systems, Man, and Cybernetics SMC-7*, 117–121.

Stefani, R. T. (1977b). Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics SMC-7*, 117–121.

Stekler, H. O. and A. Klein (2012). Predicting the outcomes of NCAA basketball championship games. *Journal of Quantitative Analysis in Sports 8*(1), Article: 1.

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

Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting 30*, 934–943.

Štrumbelj, E. and P. Vračar (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. *International Journal of Forecasting 28*, 532–542.

Teramoto, M. and C. L. Cross (2010). Relative importance of performance factors in winning NBA games in regular season versus playoffs. *Journal of Quantitative Analysis in Sports 6*(3), Article: 2.

Timmermann, A. (2006). Forecast combinations. In *Handbook of Economic Forecasting*. Elsevier Press.