Time-varying predictability of sports games

Jesse H. Krijthe

Bachelor thesis - 2 July 2010 - Erasmus University Rotterdam

Abstract

Enjoyment of sports competition is likely best when predictability is lowest. Several approaches have been developed to measure competitiveness. In this paper I discuss the amount of information that recent results carry on the outcomes of future games using the predictability standard defined by Diebold & Mariano (2001). To forecast the outcome of games I use a probit model with autoregressive terms. The approach is applied to two competitions: US Major League Baseball and English Premier league football. The latter seems to have higher predictability compared to the former. However, interpretation of the results is difficult because of the difficulties in making optimal forecasts, large variability between teams and lack of a distribution for the predictability statistic. Overal, using the predictability statistic with autoregressive probit models is not an ideal approach for measuring competitiveness in general, although a possible improvement is suggested.

Part I. Introduction

Sports competitions capture the interest of millions of people every day. Earlier research by Neale (1964) suggests that lack of competition has a negative effect on demand for a sport. He refers to this as the "League Standing Effect". Therefore a measure of predictability, derived from the predictability statistic introduced by Diebold & Kilian (2001) (DK-statistic) may help to analyze the difference in competitiveness between teams and between competitions. In particular, what this statistic would allow us to do is study the effect of recent information on our ability to predict future outcomes. If we could analyze predictability in a better way it allow sports authorities to evaluate whether rule changes have made their leagues more competitive. Additionally, this predictability measure can also be of interest to the large sports betting market, as it gives an idea whether to change their forecasts in light of recent performance of a team.

Different approaches have been used to tackle the issue of competitiveness in sports. A prevailing measure is the Competitive Balance Ratio (CBR) introduced by Humphreys (2002). The CBR is the ratio of the average variation in won-loss percentage of teams over the years divided by the average variation over the years in won-loss percentages. The idea behind this metric is that competitions with big win-loss variation over time (so different teams doing well in different years) get a high rating while competitions without this feature get low ratings. Other metrics, such as Eckhard's (1998) variance decomposition behave similarly, but have stricter assumptions.

Another approach is presented in a paper by Ben-Naim, Vazquez and Redner (2006) and considers the predictability of sports competitions by looking at parity. They consider the chance of the "underdog" winning a game. In leagues where the underdog has a big chance of upsetting the expected outcome, predictability is lowest.

In this paper I would like to add a new approach to the literature on the question of predictability in sports competitions by considering the predictability statistic introduced by Diebold and Kilian (2001), and used by de Bruijn, Bulthuis and Krijthe (2010) to explore the predictability of macro-economic variables. This new approach will mostly help to analyze the effect of recent information on the ability to forecast results. Traditional metrics discussed before take into account the amount of variation in team performance between seasons but they fail to take into account patterns during shorter periods. For example, in some motor sports it is already apparent after a few races what teams have the best technology which will not likely change during the season. Also, injuries or "winning streaks" might have more effect on some teams/competitions than others.

A large variety of approaches has been applied to the problem of forecasting sports games. They range from artificial neural nets (Wilson 1995) to Bayesian models (Yang & Swartz 2003) and simple linear models (Wood 1992 & Harville 1980). Most note the difficulty in predicting the outcomes of games (especially for baseball). As they mostly consider aggregate outcomes of competitions (such as winner of a competition or win percentage at the end of the season), single games are likely even harder to forecast correctly. In this paper a method of forecasting is used based on the work of Kauppi & Saikkonen (2001) who tried to predict US recessions using an auto-regressive probit model. To the author's knowledge this type of model has not been used on sports data before. It allows us to focus on the effect of the history of the time series on future predictions instead of an extensive list of external variables.

Using these models I will try to find whether there is a difference in the value of recent information in forecasting between different teams and between competitions. The two competitions under consideration are English Premier League football and US Major League Baseball.

Several issues will be addressed in this paper. First of all the DK-statistic has to be adapted for the case of binary outcomes in the predicted variable. Games are either predicted correctly or incorrectly. Diebold & Mariano's definition of the statistic is sufficiently general to be adapted to this case but a suitable implementation has to be defined.

Secondly, the model used for the forecasts has to be selected. In this paper I focus on the effect of recent results of games on future results. Therefore, the models used are mostly univariate. Mostly, because some variables will be present that do not follow directly from the results series of the team, team A, under consideration. For example, the recent results of the opponent team B or the recent results in games between A and B are also considered, while not directly available in the time-series for team A.

Thirdly, I will discuss the pattern that emerges for the different teams and the different sports. As I shall point out in the discussion of the data, the win percentage of teams in the baseball data are much closer together than in the football data. Therefore, in baseball there are no teams that are a "sure win" for other teams, most games can be won by either side. If recent games would have a large positive effect on future outcomes we would expect to see teams with a lot of wins and those with a lot of losses. We do see this more prominently in the football data. I therefore expect the recent games to contain more information to predict future games in football than in baseball. The opposite effect, recent games having a negative effect on future outcomes might also be possible, but is more difficult to explain.

At this point, it is also important to note what I will not discuss in this paper. This paper focuses on the information available in the win-loss series and how much this information improves forecasts. Therefore, some variables will not be considered that would be useful in forecasting. Take, for example, player salaries or their value on the transfer market. These variables might allow us to predict the outcome of games very well, since good players often have higher wages than other players. However, the information is available in both the last game of the season as well as the first game of the season. Given that these types of variables will both improve the short-run as well as the long-run forecasts, they should not affect predictability, which is a function of short-term forecast error and long-term forecast error. I will come back to this example when discussing the data.

The rest of this paper is structured as follows. Part II discusses the method used to determine the predictability statistic, generalized by Diebold & Mariano (2001), and discusses the type of model used. Part III discusses the sources of the data used in the estimation and their characteristics. Part IV contains the results of the estimation of the predictability statistic for two sports competitions. Finally, Part V concludes and discusses future topics of interest.

Part II. Methods

In this paper I consider binary won-loss time-series of sports teams y_t. Here $y_t \epsilon \{0,1\}$ denotes the result for game t in the series, where 1 means a win and 0 a loss or draw. In order to determine the value of the information at time t-j to predict the outcome at y_t we first have to define the predictability statistic. This is done below. After this I will discuss the models used to create the forecasts.

Predictability

Diebold & Mariano (2001) define a general statistic for measuring predictability. It is given by

$$P(L, \Omega, j, k) = 1 - \frac{E(L(e_{t+j,t}))}{E(L(e_{t+k,t}))} \tag{II.1}$$

Where Ω is the information set, L the loss function, j the short-term horizon, k the long-term horizon and $e_{t+h,t}$ the prediction error when forecasting y_{t+h} at time t. This statistic is broadly applicable to any covariance or difference stationary time-series. Theoretically its value is between 0 and 1, although as can be seen in de Bruijn et. al. (2010) in practice estimation might cause the statistic to drop slightly below 0. It can be interpreted as the percentage that the error of the short term forecast at time t-j is smaller than the error of the long-term forecast at time t-k. As such it measures the value of the information that is available at time t-j and not yet at time t-k on the forecast of y_t. In practice we need to define a suitable loss function L(.). The most obvious alternatives are a quadratic loss function $L(x) = x^2$ or an absolute loss function $L(x) = |x|$. How we define the error matters when choosing the loss function. Suppose we choose $e_{t+h,t} = I(p_{t+h,t} > c) - y_{t+h}$ where $p_{t+h,t}$ is the chance of a win at time t+h predicted at time t and c is the value at which we actually predict it as a win (usually this will be 0.5 unless we want to be very sure not to incorrectly pick the team as a winner). In this case both loss functions become equivalent and selection would be unnecessary. However, this would make the results dependent on the value chosen for c. Also it does not allow errors to differ in severity: forecasting a 0.8 chance of a win when there is a loss is just bad as forecasting a 0.6 chance. Therefore I use $e_{t+h,t} = p_{t+h,t} - y_{t+h}$. In this case the loss functions do differ. The quadratic function gives a bigger weight to forecasts further away from the actual value than the absolute function. This makes sense here as small errors matter less than big ones: a prediction of 0.2 chance of winning while there is a loss still gives the correct prediction for most values of c while a 0.8 chance does not. Therefore the quadratic loss function is preferable in this case.

Generally we can not calculate the DK predictability statistic directly. We therefore have to estimate it using generated forecasts. Combining all these results gives the following predictability statistic that I apply in this paper:

$$P(j,k) = \frac{(\sum (p_{t+j,t} - y_{t+j})^2)/n}{(\sum (p_{t+h,t} - y_{t+h})^2)/n} \tag{II.2}$$

Binary probit models

In order to forecast the chance of a win at time t+h, $p_{t+h,t}$, I will use a probit model without and with auto-regressive terms here. This is most similar to ARMA models usually considered when using Diebold and Kilian's predictability statistic. The model used here is taken from the description by Kauppi & Saikkonen (2006).

The probit model has the following form:

$$p_t = \Phi(\pi_t) \tag{II.3}$$

Where $\Phi(.)$ is the cumulative standard normal distribution function and π_t the 'generation mechanism' used. This generation mechanism is the actual model used containing the parameters and explanatory variables. The most obvious choice for the generation mechanism is one where only external variables are incorporated. This is given below.

$$\pi_t = \omega + x'_{t-1}\beta \tag{II.4}$$

In this paper the vector x'_{t-1} will contain the information on the recent performance of the opponent faced as time t as well as information on the outcome of previous encounters of the two teams. Other information that could be considered are other statistics relevant to performance in the sport, such as batting averages in baseball or goals in football. These types of game specific variables are very specific to the sport under consideration and require intricate knowledge of the sport to be selected. Also, in this paper I want to compare the effect of recent information between two sports, English football and American baseball. Therefore these variables are not used here.

The former specification does not allow us to model auto-regressive behaviour of the series. We therefore extend it to the following form.

$$\pi_t = \omega + \sum_{i=1}^{l} \alpha_i \pi_{t-i} + x_{t-k}^{'} \beta \tag{II.5}$$

This specification allows for the inclusion of l (for which we use 1 here) autoregressive terms as well as external variables in the vector x. Note that x_{t-k}' is not the lagged value of x but the value of x at time t-k. For example, if x is the average performance of the opponent, we want the most recent value available at time t-k about the opponent at time t, not the opponent at time t-k!

The parameters estimates of this model can be obtained using Maximum likelihood estimation. Take $\theta = [\omega, \alpha', \beta']$' we find their estimates by maximizing the log likelihood function:

$$L(\theta) = \sum_{t=1}^{T} [y_t log(\Phi(\pi_t(\theta))) + (1 - y_t) log(\Phi(\pi_t(\theta)))] \tag{II.6}$$

The next step in the process is to make the forecasts. For the first model (equation II.4), the static model, I will use a moving window as well as horizon-specific models to generate forecasts. This means to forecast observation y_{t+1} the model is estimated using the N most recent observations available at the horizon under consideration. For instance, when the horizon is h games, observations $y_{t-h-N}, y_{t-h-(N-1)}, ..., y_{t-h}$ will be used to estimate the model parameters. This is done for all $y_{t+1}, y_{t+2}, ..., y_{t+M}$ where M is the length of the forecast sample in the forecast sample. This means the model will be reestimated M*H times, where H is the maximum horizon.

Although this is a computationally intensive task, the alternative is getting worse forecasts at long-horizons than possible. This would intentionally reduce the accuracy of the forecasts in the denominator of equation II.2, generating too large values of the predictability statistic. We might unjustly conclude predictability is high while in fact, it is artificially created to be high by generating bad long-term forecasts.

On the downside, reestimating the model at different horizons might occasionaly cause the forecasts to be more accurate at longer horizons due to parameter estimation. In other words: the predictability graphs will not be as smooth as they would be in theory and predictability values might become negative. Although this makes it harder to interpret the results, I believe the advantage stated earlier outweights these difficulties.

In the case of the autoregressive model (equation II.5), the computations using this approach turned out to be too time consuming. Therefore, direct forecasts have been calculated while still using a moving window. The direct form of the forecasts is as follows:

$$E_{t-h}(y_t) = \Phi(\alpha^h \pi_{t-h} + \sum_{j=1}^{h} \alpha^{j-1}(\omega + \beta x_{t-k+1-j})) \tag{II.7}$$

Where k is chosen equal to the maximum horizon under consideration. For the derivation of these direct forecasts I refer to Kauppi & Saikkonen (2006).

Part III. Data

The methods described in the previous section will be applied to two sports competitions, US Major League Baseball (MLB) and English Premier League Football. Both have different data sources and characteristics which will be discussed below.

Football

The football data has been taken from Football-data.co.uk[1] which in turn is a collection of data from several major sports websites. The results from the 2008-2009 season are given in Table 1[2]. Total team wages and debt have been provided for all teams (except Liverpool) as well. In the estimation of the predictability statistic in the next section, data from the two previous seasons is used for parameter estimation as well (season 2006/2007 and 2007/2008).

Characteristic of the data is the large difference in winning percentage between the team ranking first and the team ranking last: 0.526. This might suggest there is a large difference in the quality of the teams. Previous win percentage could

[1] see: http://www.football-data.co.uk/englandm.php

[2] Results from http://www.fcupdate.nl/stand/s259/engeland-premier-league-2008-2009/ and wages and debts from The Guardian through http://www.thevillablog.co.uk/aston-villa-blog/avfc-news/aston-villa-come-in-ninth-in-the-premier-league-list-of-clubs-by-turnover/

Team	Wages (£M)	Debt (£M)	Ranking	Won %
Manchester United	252.6	121.1	1	0.737
Liverpool	159.0	NA	2	0.658
Chelsea	213.6	149.0	3	0.658
Arsenal	222.5	101.3	4	0.526
Everton	76.0	44.5	5	0.447
Aston Villa	75.6	50.4	6	0.447
Fulham	53.7	39.4	7	0.368
Tottenham	114.7	52.9	8	0.368
West Ham	57.0	44.2	9	0.368
Manchester City	82.3	54.2	10	0.395
Wigan	43.0	38.4	11	0.316
Stoke	11.2	11.9	12	0.316
Bolton	59.1	39.0	13	0.289
Portsmouth	70.5	54.7	14	0.263
Blackburn	56.4	39.7	15	0.263
Sunderland	62.6	37.1	16	0.237
Hull	9.0	6.9	17	0.211
Newcastle	100.6	74.6	18	0.184
Middlesbrough	48	34.8	19	0.184
West Bromwich	27.2	21.8	20	0.211

Tab. 1: Results from the 2008/2009 Premier League season including wages and debt

therefore predict the outcome of the game. For example, the team currently with the high win percentage will likely beat the team with the low win percentage. Note that in football it is possible to rank higher and still have a lower fraction games won due to the possibility of draw games. This is the case for Manchester City, Newcastle and Middlesbrough.

I described earlier that some variables might help predict a winner but are not considered in this paper due to the information being available in both the sparse (long-term) and the rich (short-term) information set. One example is the wages of the players in a team or the total debt (which, in football, often seems to serve as a proxy for investment in good players). Both these variables have high correlation with the actual result: -0.72 and -0.65 respectively. The correlation of won percentage and wages is even greater at 0.84. This gives an indication that these variables might be useful in predicting the future outcome of the Premier league. However, for the reasons stated earlier, they are not relevant in the analysis in this paper.

Unlike the Major League Baseball, the teams that make up this competition change every year. Each year the teams ending last are relegated to a different league and other teams are added. For this reason, there is more information on the performance of some teams than others. In the results I will only discuss teams that have had a presence in the Premier league over all the years in the sample. In most seasons this approach would not allow me to study the lowest ranking teams. However, in the season under consideration long term premier league contenders, Middlesbrough as well as Newcastle, performed among the worst teams in the league. Therefore we can study the effect on the teams at the bottom of the league. A problem that remains is that some opponents of the teams analyzed have not had a long history in the league, and therefore have few information on past performance. Since we can not just leave out some values out of the time series I use values that represent a no-information guess. For example, if a team has not encountered an opponent before, I assume we can not say he will perform better or worse than any other opponent. It could be argued this is too optimistic: opponents without previous encounters are likely recently promoted to the league and will therefore perform worse than the average opponent.

A last characteristic is that football matches can end in a draw. To simplify the model and be able to compare the outcomes between two sports I only consider the difference between win and non-win (draw or loss).

Baseball

MLB data was collected from the Retrosheet project using some tools available from the very active Sabermetrics (baseball statistics) community[3]. Statistics go back to before 1900, but here only recent years (2008 for the sample and 2009 for the forecasts) have been included. In recent years the MLB has had 30 teams

[3] See retrosheet.org. Data collection was greatly simplified by the tools provided by the Chadwick project and code published by Wells Oliver see: http://blog.wellsoliver.com/2009/06/retrosheet/

competing, divided over 6 divisions. The results[4] for one division in the 2009 regular-season are given in Table 2.

Team	Wages($M)	Ranking	Won%
New York Yankees	201.4	1	0.636
Boston Red Sox	121.7	2	0.586
Tampa Bay Rays	63.3	3	0.519
Toronto Blue Jays	97.8	4	0.463
Baltimore Orioles	80.5	5	0.395

Tab. 2: Results from the 2009 MLB season for the American League East

The baseball data are somewhat different from the football data. As can be seen in the table, the difference in win percentage between the best and worst team is smaller here (only 0.241). In the total league the difference is 0.272. Still this is a lot smaller than in the case of the Premier league. This might suggest it is much harder to predict the outcome of a match based on the previous win percentages, because the percentages of both teams are very close to the 'no-information' case of 0.50/0.50.

Correlation between wages and won percentage is also much smaller here than in the Premier league at 0.45 compared to 0.84.

The amount of games played in the MLB per season is much larger than in the English Premier League: 162 per team instead of 38. During this time, we can assume team rosters are relatively stable. This is an advantage while estimating the parameters, because it is more likely parameters stay the same during a season than between seasons.

Also the league itself is very stable: there are no relegations or promotions. In the period discussed in this paper, no changes were made to the league.

Part IV. Results

I will first discuss the results for the English Premier league competition. Then I will use the same methods to analyze the Major League Baseball competition and

[4] Results from http://sports.yahoo.com/mlb/standings?type=regular&year=season_2009 and wages from http://baseball.about.com/od/newsrumors/a/09teamsalaries.htm

finish with a short comparison of the two.[5]

English Premier League Football

The estimation results are shown in Figure IV.1. The left panel shows the static probit model from the specification in II.4. There are two variables used here: the average number of wins in the last 10 available games of the opponent and the average number of wins by this team against the opponent. Both variables are not always available. For instance some opponents just spend one year in the Premier league, so there have not been enough matches to determine these variables. In these cases the average number of wins is taken as 1/3 (because we have no information than the fact that any game can be a either a won, a loss or a draw). For the performance in recent encounters I take the value of the average number of recent games won by the team under consideration, because we have no way of knowing whether it will perform better or worse against this opponent. However, these replacements do not have to be used very often since most teams have played enough games in the past.

Other variables that were considered but did not positively effect the outcome of the predictability values were previous values in the won loss series. The binary nature of these outcomes might make it harder to use these as predictors of future performance: a loss does not neccesarily mean a team performed badly. Therefore I also used the average performance in the last 5 or 10 games as a variable. The parameter estimate for this variable does not significantly differ from zero in most of the regressions. This could be caused by the reestimation procedure as the intercept $\hat{\omega}$ of the equation is constantly updated with information of recent outcomes. Therefore the variable contains the same information as the intercept, only over a smaller time period. If very recent performance did help predict future outcomes, we would expect this variable to perform well. However, this does not seem to be the case.

As can be seen from the figure, some of the graphs behave quite irratically. This is especially true if the other teams not included in the figure (for reasons of

[5] To estimate the models, the R functions glm, for the static probit models, and optim, using an appropriate likelihood function for the autoregressive probit model were used

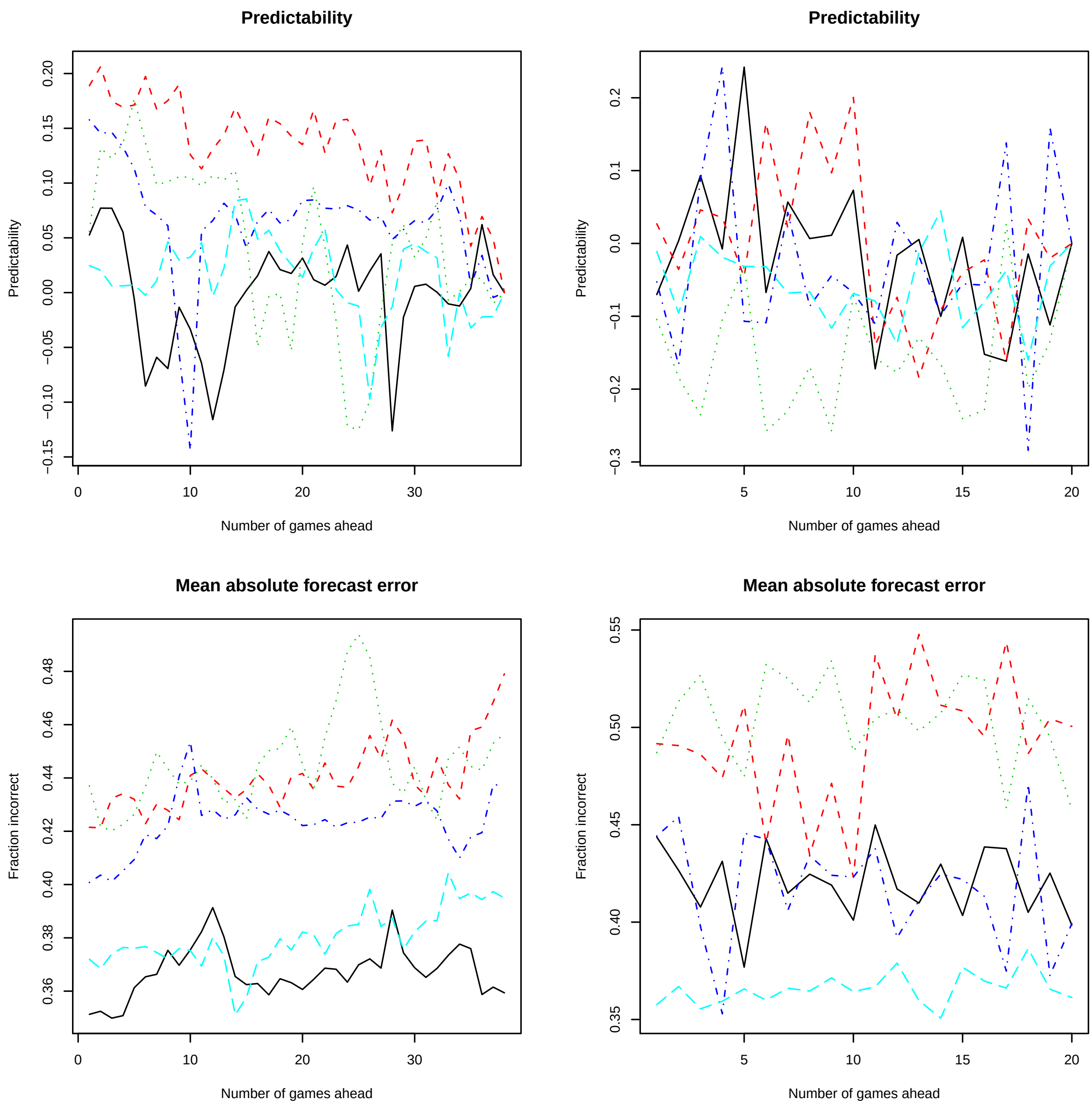

Fig. IV.1: Predictability estimation results for English Premier League football using the static probit model (left) and the autoregressive probit model (right) for selected teams. The top figures are the predictability outcomes, the bottom figures the fraction of games incorrectly predicted (win, non-win) when using a cutoff value of c=0.5. Teams are as follows (in top left figure from top to bottom at horizon 1): Everton, Bolton, Aston Villa, Manchester United (solid line) and Middlesbrough.

clarity) are also considered. This shows the difficulty in estimating the parameters and the problem of this model: sometimes longer-horizon forecasts work better than shorter-horizon forecasts. Theoretically this should not be possible if all information where used optimally. However, the models used here do not adhere to this optimality condition. Because of this, it is hard to determine what effects are actual effects of the value of the information to improve the forecast, and what part is noise. This is further complicated by the lack of a known distribution for the predictability statistic.

The second model, shown on the right, is the autoregressive probit model from the specification in II.5. In this case I have selected just one of the variables to include in the vector x: the average number of games won by the opponent in its last 10 games. The results in this case are less clear than in the static model. There are many values below zero and the fraction predicted incorrectly is mostly larger than in the static model. This might be improved if more variables are added to the equation. However, the results also do not fully agree with those obtained from the static model: Everton seems to still perform well while Manchester performs better and Aston Villa performs worse.

Comparison teams

Both of the panels in the figure contain the outcome of selection of five teams to keep the figures from cluttering: Manchester United (ranked 1st), Everton (5th), Aston Villa (6th), Bolton (13th) and Middlesbrough (19th, lowest winning percentage). From the static model it seems Everton performs best followed by Bolton (except for the sudden drop at horizon 10). Recent information also seems to have a positive effect on Aston Villa but the effects on Manchester and Middlesbrough are too close to zero to not attribute them to the randomness in the graphs. In the dynamic model we see the same effect for Everton and Bolton, however the results are much more erratic and it is hard to draw any conclusion from this.

An interesting outcome is the effect of the variables on the fraction incorrectly predicted. In the static model for most teams adding these variables adds 5% to the percentage we predict correctly compared to the static model with only an intercept. For most teams this increase happens at every horizon. The variable serves as a measure of the opponent's time independent strength, while in the case of the teams where there is only an increase at shorter horizons the strength of the opponent at a certain point in time seems more important.

In general, I find no relationship between placement in the league and predictability. Teams that do perform well do seem to be situated in the middle of the league table with winning fraction closer to 0.5 (for example Everton and Aston Villa both ended with 0.447), but there are quite a few exceptions to this.

Major League Baseball

A similar method was used for the MLB data. The results have been provided in Figure IV.2.

For the static model, the same variables where used here as in the football dataset. The horizon chosen here is much smaller than before because this greatly reduced the time to compute the results. The results for the static model look promising with few values below zero. However, this is certainly not true if we consider all teams (more results have been provided in the appendix).

The dynamic model yields very different results: the values are mostly negative. Some results agree with the static model, for example Philadelphia and New York while others (Washington) show the opposite effect.

Interestingly, if we compare the fraction incorrectly predicted for both models the autoregressive model gives results that for some teams are much better (New York decreases 5%) and others much worse (Washington increases 5%) than the static model. Clearly the autoregressive model works better for some teams than others.

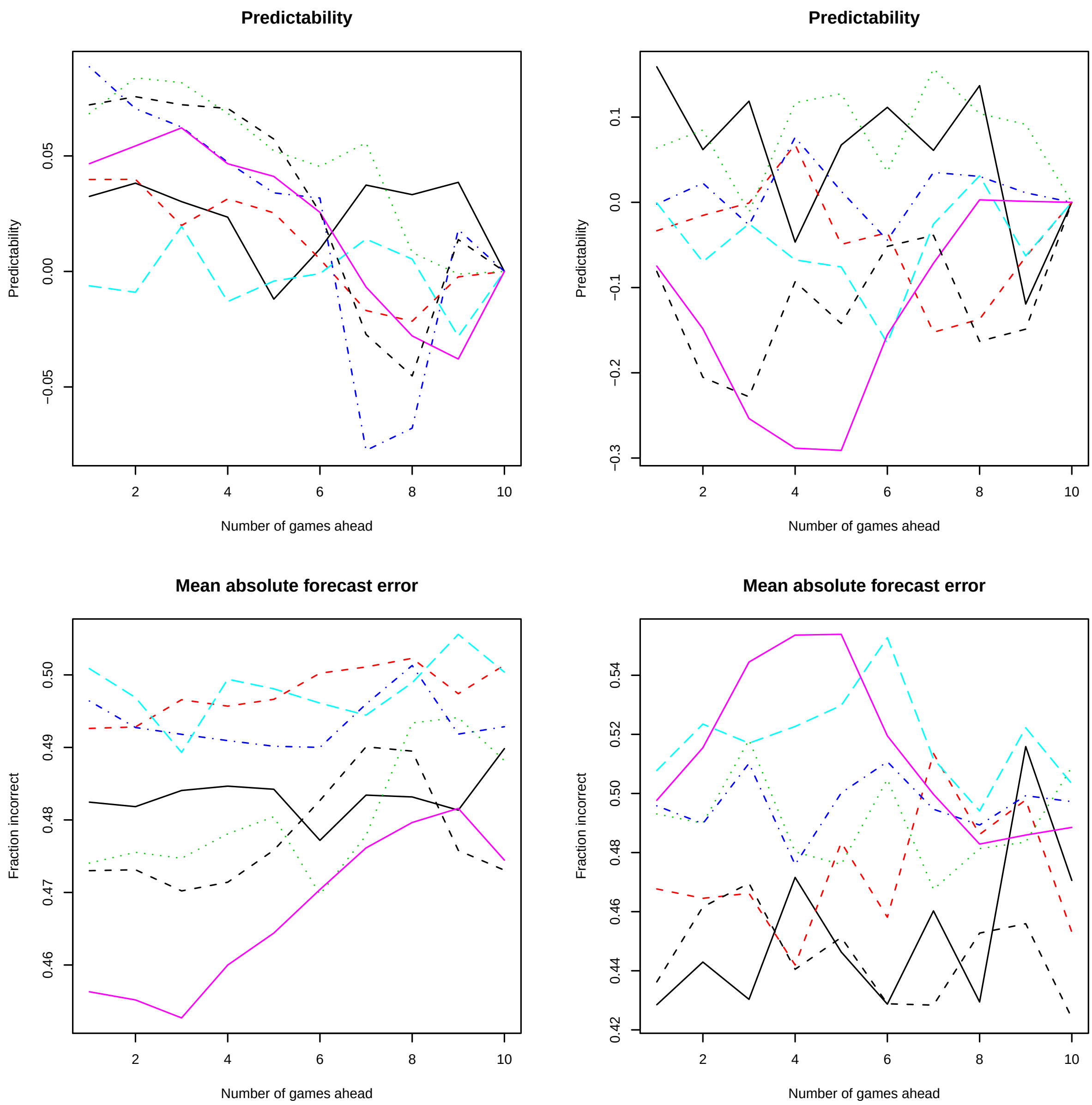

Fig. IV.2: Predictability estimation results for Major League Baseball using the static probit model (left) and the autoregressive probit model (right) for selected teams. The top figures are the predictability outcomes, the bottom figures the fraction of games incorrectly predicted when using a cutoff value of c=0.5. Teams are as follows (in top left figure from top to bottom at horizon 1): St. Louis, Pittsburgh, Philadelphia, Washington, LA Dodgers, NY Yankees and Chicago White Sox.

Comparison teams

The teams displayed in Figure IV.2 are (with end-year ranking and winning percentage): St. Louis (2nd, 0.544), Pittsburgh (6th, 0.384), Philadelphia (3rd, 0.532), Washington(5th, 0.438), LA Dodgers (2nd, 0.551), NY Yankees (1st, 0.615) and Chicago White Sox (3rd, 0.519). There does not seem to be a pattern in the results. For example, Philadelphia and Chicago performed almost as well while their predictability results are very different.

Comparison sports

The high variability in the predictability estimates makes it hard to compare the two sports. One feature that does stand out is that the values in the football competition are larger in general, than in the baseball data. Also the fraction of games forecasted incorrectly tends to be smaller in football. It seems to be easier to get reasonable results in football than in baseball. This is mostly a feature of the type of distribution of wins in the sport, in football predicting a win every game for the 'big four' (Manchester United, Arsenal, Liverpool, Chelsea) and a loss every game for the other games already gives better results than in baseball (where the win percentages are closer to 0.500).

Part V. Concluding remarks

In this paper I a predictability measure with probit models to analyze the effect of recent results on our ability to predict future results in football and baseball competitions. The results are not unambiguous and the method used turns out not to be very effective in answering the question of predictability and competitiveness.

In baseball recent results do not seem to contain much information on future games. This finding agrees with the analysis by Wood (1992) who find for baseball games that “outcome of the previous game did not account for a significant amount of the variance in the outcome of individual games ”. In this paper variance explained is not considered directly, but low variance explained by these variables would also cause only small improvements in predictability, if any. However, the results presented here are not unambiguous. The erratic behaviour of the predictability statistic suggests it is difficult to incorporate all information available at a given horizon. The results could therefore lead to two conclusions: either the games are not predictable or other variables are needed in the model that are available at time t-h.

The football data shows a similar pattern, although some teams do have a higher predictability than others. There seems to be a pattern that teams in the middle of the rankings with win percentages near 0.500 have higher predictability at lower horizons. This could be because for these teams it is hard to make a prediction and the extra information has a large impact on getting a better estimate than a 50/50 chance. The marginal effect of this extra information on a team who wins (or loses) almost every game is likely much smaller.

Comparing the two sports, football seems to have somewhat higher predictability values as well as lower percentages of incorrectly predicted games. In this sense there might to be some evidence for the hypothesis in the introduction that football has higher predictability because of the bigger difference in win percentages of the teams. However, the lack of a distribution for the predictability statistic as well as the difficulties in using all the information in the models makes this a very weak claim. It simply gives an indication of a result, unfortunately, nothing more.

In the previous section we have seen the difficulty of getting stable and clear-cut estimates for the predictability statistic. To get good estimates for the predictability statistic the data has to be used optimally at every horizon to generate forecasts. As can be seen from the many negative values in the results, the models used here clearly do not pocess this property.

The models used here do offer an improvement over the "base model" which is a static probit model with only an intercept. The autoregressive model in particular gives better results for some teams. However, this last model suffers even more from the estimation problems mentioned earlier. It does tells us that the models might be a starting point to the creation of a model to predict these games, although better alternatives seem to be available in the literature.

Overall, the predictability statistic with the static and autoregressive probit model turn out not to be a good way to measure competitiveness in general for two reasons. First of all, it is very difficult to create forecasts that make optimal use of all information, making it hard to estimate the statistic. Secondly, the recent information that the outcome of sports games dependent on might be more adequately captured in other forms than recent results. Some would suggest variables such as player salary or market value, but they are not applicable here. These are long-term variables (long-term when looking at predicting the next game) and so they would be available at tie t-j as well as t-k. Other variables that do change over the period is news on injuries or player well-being or the general tendency of media to favour a certain team.

One way to overcome both difficulties is modelling this information using betting odds. One possible approach would be to record the betting odds over time. If the betting market is efficient and all information is correctly incorporated in the odds, this would be a good proxy of a prediction containing all information available at time t-j. We could then evaluate how well the game was predicted at time t-j and use these forecast errors to calculate the predictability statistic. The problem with this approach might be that most betting lines only open shortly before the game, although for some important games this might be different. It is also very questionable whether the sports betting market is efficient and incorporates all information, as has been discussed by Woodland & Woodland (1994) and Gray & Gray (1997) among others. However, the information might still be incorporated enough to help shed light on the question of predictability over time in sports results.

References

[1] Ben-Naim, Eli; Vazquez, Federico; and Redner, Sidney (2006) "Parity and Predictability of Competitions", Journal of Quantitative Analysis in Sports: Vol. 2 : Iss. 4, Article 1.

[2] De Bruijn, Bert; Bulthuis, Fleur; Krijthe, Jesse (2010) "Predictability of macro-economic variables", Werkcollege Algemene Econometrie, Erasmus University Rotterdam

[3] Diebold, Francis X.; Kilian, Lutz (2001) "Measuring Predictability: Theory

and Macroeconomic Application", Journal of Applied Econometrics, Vol. 16, No. 6 (Nov. - Dec., 2001), pp. 657-669, John Wiley & Sons

[4] Gray, Philip K.; Gray, Stephen F. (1997) "Testing Market Efficiency: Evidence from the NFL Sports Betting Market" The Journal of Finance, Vol. 52, No. 4 (Sep., 1997), pp. 1725-1737

[5] Harville, David (1980) "Predictions for National Football League Games Via Linear-Model Methodology", Journal of the American Statistical Association, Vol. 75, No. 371 (Sep., 1980), pp. 516-524

[6] Humphreys, Brad R. (2002) "Alternative Measures of Competitive Balance in Sports Leagues", Journal of Sports Economics, Vol. 3, No. 2, 133-148 (2002)

[7] Kauppi, H.; Saikkonen, P. (2006) "Predicting U.S. Recessions with Dynamic Binary Response Models", Helsinki Center of Economic Research Discussion Paper No. 79.

[8] Neale, W. C. (1964) "The peculiar economics of professional sports", Quarterly Journal of Economics, 78, 1-14.

[9] Rue, H.; Salvesen, Ø. (2000) "Prediction and retrospective analysis of soccer matches in a league". Statistician, 49, 399–418.

[10] Wilson, Rick L. (1995) "Ranking College Football Teams: A Neural Network Approach", Interfaces, Vol. 25, No. 4 (Jul. - Aug., 1995), pp. 44-59

[11] Wood, G. (1992) "Predicting Outcomes: Sports and Stocks", Journal of GamblingStudies Vol. 8(2), Summer 1992

[12] Woodland, Linda M.; Woodland, Bill M. (1994) "Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market" The Journal of Finance, Vol. 49, No. 1 (Mar., 1994), pp. 269-279

[13] Yang, Tae Young; Swartz, Tim (2003) "A Two-Stage Bayesian Model for Predicting Winners in Major League Baseball ", Journal of Data Science 2(2004), 61-73

Appendix

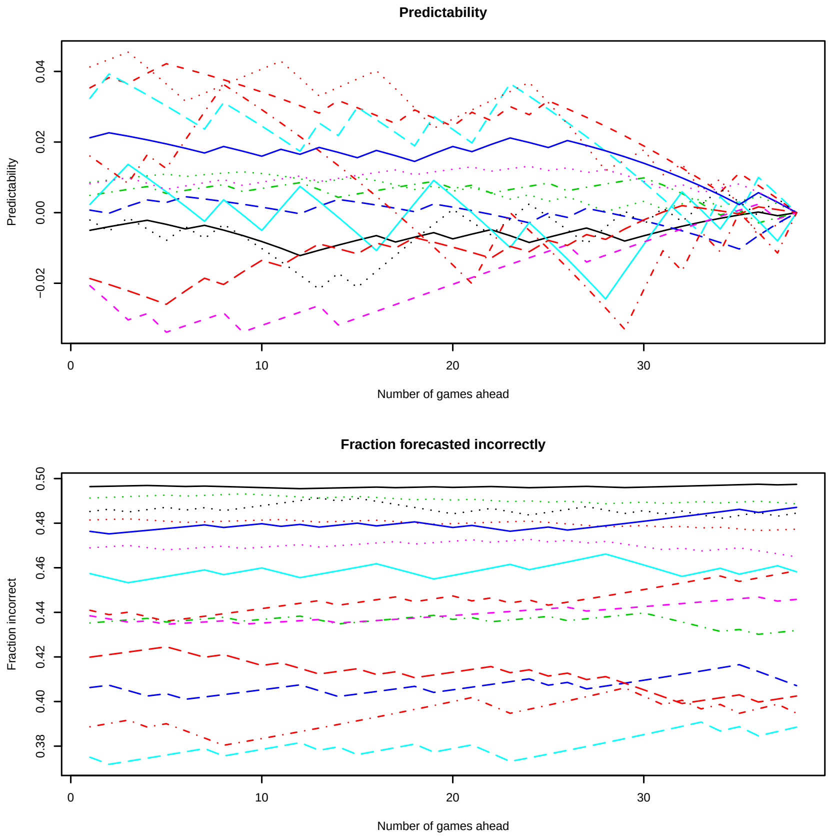

Fig. V.1: Top: Predictability for all teams in the football dataset when using the static model with only an intercept. Bottom: the corresponding fraction of correct observations. This serves as a baseline value for the fraction of incorrectly forecasted games.

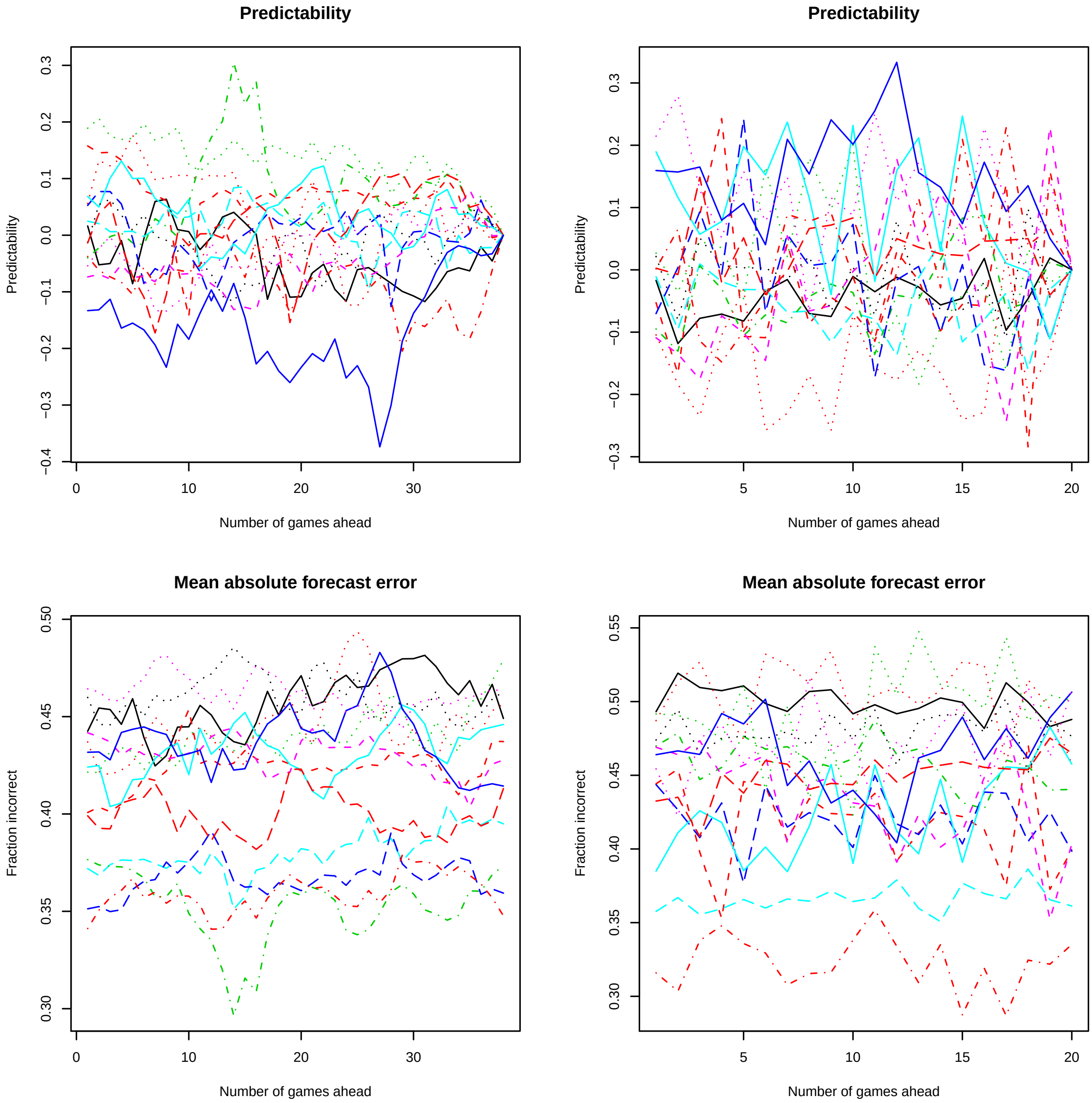

Fig. V.2: Predictability estimation results for English Premier League football using the static probit model (left) and the autoregressive probit model (right) for all teams. The top figures are the predictability outcomes, the bottom figures the fraction of games incorrectly predicted (win, non-win) when using a cutoff value of c=0.5

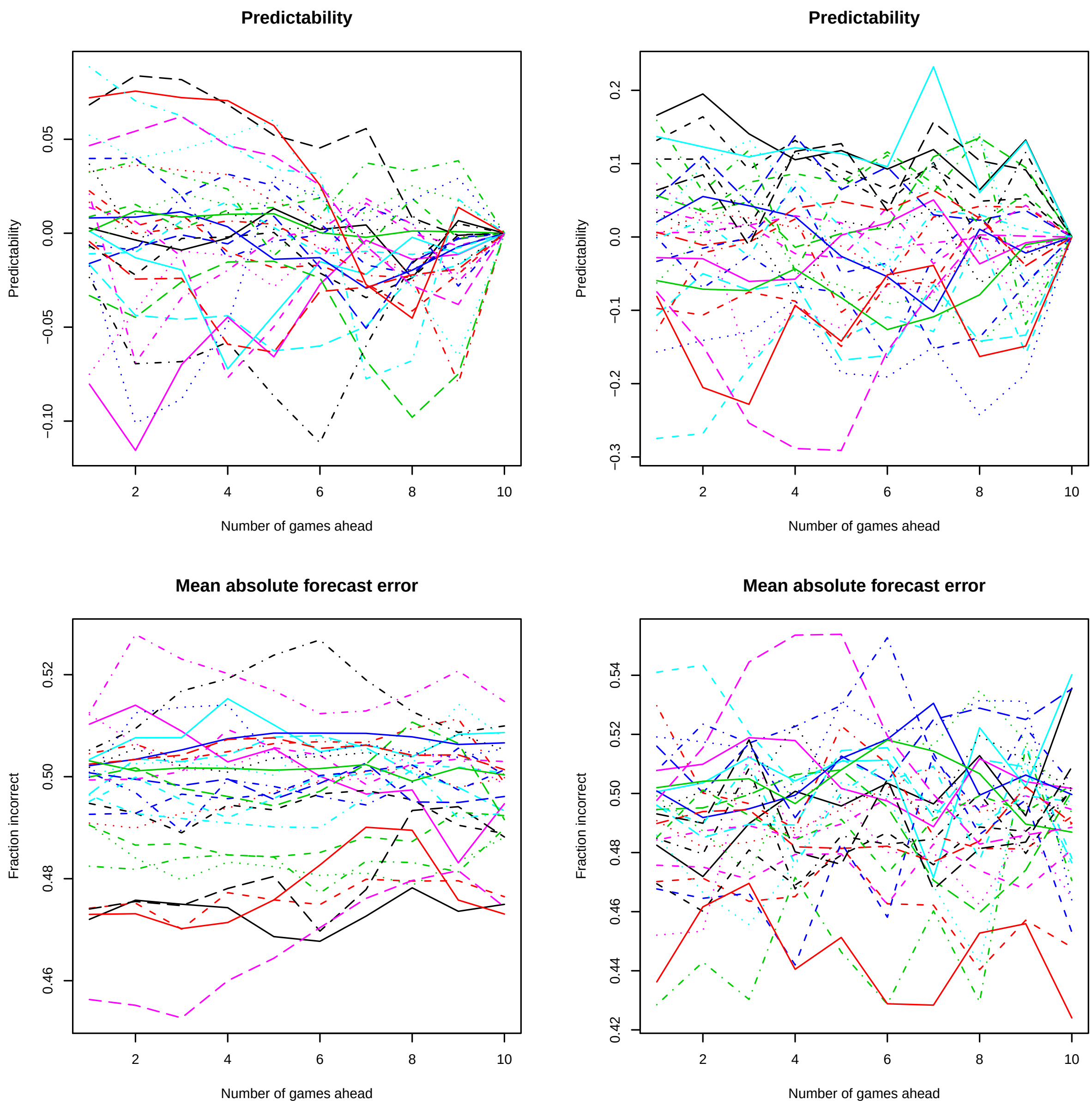

Fig. V.3: Predictability estimation results for Major League Baseball using the static probit model (left) and the autoregressive probit model (right) for all teams. The top figures are the predictability outcomes, the bottom figures the fraction of games incorrectly predicted when using a cutoff value of c=0.5