Christoph Leitner and Achim Zeileis and Kurt Hornik

Bookmaker Consensus and Agreement for the UEFA Champions League 2008/09

Working Paper

Bookmaker Consensus and Agreement for the UEFA Champions League 2008/09

Christoph Leitner, Achim Zeileis, Kurt Hornik

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 88
August 2009

http://statmath.wu-wien.ac.at/

Bookmaker Consensus and Agreement for the UEFA Champions League 2008/09

Christoph Leitner*, Achim Zeileis*, Kurt Hornik*

* Department of Statistics and Mathematics, WU Wirtschaftsuniversität Wien, Augasse 2–6, 1090 Wien, Austria

Abstract

Bookmakers odds are an easily available source of "prospective" information that is thus often employed for forecasting the outcome of sports events. To investigate the statistical properties of bookmakers odds from a variety of bookmakers for a number of different potential outcomes of a sports event, a class of mixed-effects models is explored, providing information about both consensus and (dis)agreement across bookmakers. In an empirical study for the UEFA Champions League, the most prestigious football club competition in Europe, model selection yields a simple and intuitive model with team-specific means for capturing consensus and team-specific standard deviations reflecting agreement across bookmakers. The resulting consensus forecast performs well in practice, exhibiting high correlation with the actual tournament outcome. Furthermore, the teams' agreement can be shown to be strongly correlated with the predicted consensus and can thus be incorporated in a more parsimonious model for agreement while preserving the same consensus fit.

Keywords: Consensus, agreement, bookmakers odds, sports tournaments, Champions League.

1 Introduction

In the course of growing popularity of online sports betting, the analysis of betting markets has been receiving increased interest, often focusing on two types of analyses: (1) testing the forecasting power of the bookmakers, and (2) testing the efficiency of the betting market. Here, we use bookmakers odds from a variety of bookmakers as forecasting information to predict sports tournament outcomes and apply this in an empirical study for the *UEFA Champions League 2008/09*. More specifically, we employ a model that not only reflects the bookmakers' "consensus" for forecasting but also captures the "agreement" or "disagreement" across the bookmakers.

Bookmakers odds are prospective ratings of the performance of the participating players or teams in a sports competition which vary between the bookmakers. They were successfully used to predict the outcome of single games (e.g., Spann and Skiera, 2009; Song *et al.*, 2007; Forrest *et al.*, 2005; Dixon and Pope, 2004; Boulier and Stekler, 2003). Based on these ideas, Leitner *et al.* (2008) and Leitner *et al.* (2009) use aggregated quoted odds of a variety of bookmakers to forecast the outcome of whole tournaments, the EURO 2008 and the UEFA Champions League 2008/2009, respectively. Their studies performed successfully, in particular predicting the final of the EURO 2008 correctly.

To aggregate information of different forecasters a measure of "consensus" is needed. Zarnowitz and Lambros (1987) define "consensus" as the degree of agreement among point predictions aimed at the same target by different individuals and "uncertainty" as the diffuseness of the corresponding

probability distributions. It can be computed as the median (Su and Su, 1975) or the mean of all the forecasts in the sample (Zarnowitz and Lambros, 1987). To measure "uncertainty" or "disagreement" the standard deviations of the predictive probability distributions could be used (e.g, Clements, 2008; Zarnowitz and Lambros, 1987; Lahiri and Teigland, 1987). These strategies are applied to sports competitions by Song *et al.* (2009, 2007). Alternative strategies for the aggregation of forecasts are discussed by Kolb and Stekler (1996); Schnader and Stekler (1991).

Here, we follow Leitner *et al.* (2009) and extend their model framework for bookmakers odds stemming form different bookmakers to a more general model class, modeling the bookmakers consensus as well as the (dis)agreement across the bookmakers. In particular, we aggregate the quoted long-term odds of winning the UEFA Champions League 2008/09 for all 32 participating teams of 31 international bookmakers (published prior to the tournament) to a consensus measure and compute the agreement/disagreement across the bookmakers. The UEFA Champions League is the most prestigious club competition of the Union of European Football Associations (UEFA) and so one of the most popular annual sports tournaments all over the world.

A natural strategy to obtain both measures, consensus and agreement, could be the computation of the means and the standard deviations of bookmaker information for the participating teams of a tournament across the bookmakers on an appropriate scale. But is this really the best strategy? To analyze the influence of some other effects, we employ a mixed-effects model framework for the winning logits derived from the bookmakers odds capturing different effects associated with the participating teams, the bookmakers and the team's associations. This leads to a variety of mixed-effects models. After establishing the general modeling approach a subsequent model selection yields the final model including team-specific fixed effects and team-specific standard deviations. Hence, the natural idea discussed before is an appropriate approach for the computation of consensus and agreement. A high correlation between the rankings derived from the actual tournament outcome and from the bookmaker consensus confirms the good forecasting power of bookmakers. The team-specific standard deviations imply that the agreement across the bookmakers is high, where the agreement increases with increasing winning probabilities of the teams, i.e., the agreement across the bookmakers is higher for teams with high winning probabilities than for teams with low chances of winning the tournament. In addition, we analyze the relationships between the team's associations, consensus, and agreement.

The remainder of this paper is organized as follows: Section 2 provides a tournament and data description for the UEFA Champions League 2008/09 for which the bookmakers consensus and agreement are modeled in Section 3 and analyzed in Section 4. Section 5 concludes the paper.

2 UEFA Champions League 2008/09: Tournament and data description

2.1 Tournament

The UEFA Champions League is an annual tournament where a selection of European football clubs compete in a multi-stage modus (qualification, group, and knockout stage) to determine the "best" European team. First, 32 teams are determined via three qualification rounds for the group stage and drawn into eight groups (A–H). The number of eligible teams is determined by UEFA's Coefficient Ranking System for its member associations (see below, for more details). In the 2008/09 season, teams from 17 associations out of UEFA's 53 members qualified for the group stage. The four teams of each group play a round-robin—every team plays every other team twice (one home and one away match), for a total of twelve games within the group—and the group winners and runners-up advance to the knockout stages. In the knock-out stage, each round pairings are determined by means of a draw and played under the cup (knock-out) system, on a home- and-away basis, where the winners advance to the next round until two teams remain.

The two teams play the final as one single match at a neutral venue yielding the winner of the UEFA Champions League (Union of European Football Associations, 2009).

2.2 Data

Bookmakers odds. Long-term odds of winning the UEFA Champions League 2008/09 were obtained from the websites of 31 international bookmakers for all 32 participating teams on 2008-09-01 (before the tournament started, but after the group draw). The 31 bookmakers are all bookmakers who offer odds for this event out of 50 European top-selling online sports bookmakers.

The quoted odds of the bookmakers do not represent the true chances that a team will win the tournament, because they include the stake and a profit margin, better known as the "overround" on the "book" (for further details see e.g., Henery, 1999; Forrest *et al.*, 2005). To recover the underlying beliefs of the bookmakers, we have to adjust the quoted odds. We first reduce the quoted odds by one, the stake, i.e., the payment for placing the bet. Subsequently, we adjust it by the profit of the bookmaker (the overround). Assuming that the overround δ is constant across all competitors winning a tournament, it can be computed by restricting the corresponding probabilities to sum to unity. More precisely, the raw quoted odds $rawodds_i$ for team i from a single bookmaker can be adjusted to $odds_i$ and then transformed to probabilities p_i via:

$$odds_i = (rawodds_i - 1)\,\delta, \tag{1}$$

$$p_i = 1 - \frac{odds_i}{1 + odds_i}. \tag{2}$$

Then, δ can be chosen such that $\sum_i p_i = 1$. (Note, that the complementary probabilities have to be used as the bookmakers odds represent expectations for an outcome not to occur.) Therefore, the implied winning probabilities for each team can be easily derived from the quoted odds of all teams. This adjustment is done separately for all 31 bookmakers yielding bookmaker-specific overrounds δ_b and expected winning probabilities $p_{i,b}$ for each team $i = 1, \ldots, 32$ and bookmaker $b = 1, \ldots, 31$.

UEFA's club coefficient and seeding. The UEFA also announces their expectancies for the tournament outcome prior to the tournament by publishing UEFA's seeding for the group draw which is just a ranking and very similar to the ranking of UEFA's club coefficient of the teams. The UEFA's club coefficient is determined by the results of a club in European club competition in the last five seasons, and the league coefficient (for more details see Union of European Football Associations, 2009). We obtained the UEFA's club coefficient and seeding for the group draw from 2008-08-28 from UEFA's website for all 32 participating teams and employ both in comparison of the consensus forecast.

3 Modeling consensus and agreement

3.1 Model class

Based on the expected winning probabilities $p_{i,b}$ for each team $i = 1, \ldots, 32$ and bookmaker $b = 1, \ldots, 31$, as derived from the adjusted and transformed bookmakers odds, we want to obtain measures for consensus and (dis)agreement among the bookmakers. An intuitive and straightfoward strategy would be to compute this information on a logit scale, employing team-wise means for the consensus and team-wise standard deviations for the disagreement across bookmakers (as suggested by, e.g., Song *et al.*, 2009, 2007; Zarnowitz and Lambros, 1987). For our application this simple strategy might be appropriate because we could expect that the teams are sufficiently different and the bookmakers have rather similar models and information about the teams.

Nevertheless, from a statistical point of view it would be interesting to know if this simple strategy is sufficient or can be improved by including additional effects (e.g., pertaining to the bookmakers), or by using a more parsimonious parametrization which still gives a good approximation of the underlying data-generating process. Therefore, we propose a stochastic model capturing the underlying probability distribution on a logit scale. The model relates the adjusted winning logits $\text{logit}(p_{i,b})$ to the (unobservable) "true" winning logits $\text{logit}(p_i)$ for team i, reflecting the bookmakers consensus, plus an additional (unobservable) "error" term $\epsilon_{i,b}$ of bookmaker b for team i, reflecting the disagreement across the bookmakers. To capture these latent quantities by a linear mixed-effects model, we allow the true winning logits to depend on a team effect α_i, an association effect $\beta_{a(i)}$ for association a of team i, as well as an overall intercept ν. The error can additionally depend on μ_b, the effect of bookmaker b. We allow also different specifications of the deviation $\epsilon_{i,b}$ of bookmaker b for team i. In summary, this can be written as

$$\begin{aligned} \text{logit}(p_{i,b}) &= \text{logit}(p_i) + \epsilon_{i,b} && (3) \\ &= \nu + \alpha_i + \beta_{a(i)} + \mu_b + \sigma_{i,b} Z_{i,b}, && (4) \end{aligned}$$

where $Z_{i,b}$ is a standardized error and $\sigma_{i,b}$ is the standard deviation which can be either constant ($\sigma_{i,b} = \sigma$), or constant within a specific group ($\sigma_{i,b} = \sigma_i$: team-specific standard deviation, $\sigma_{i,b} = \sigma_b$: bookmaker-specific, or $\sigma_{i,b} = \sigma_{a(i)}$: association-specific). Even if contrasts are employed, this model is overspecified when all three effects α_i, $\beta_{a(i)}$, and μ_b are included as fixed effects due to the the dependence of association $a(i)$ on the team i.

To overcome this methodological issue, there are various conceivable solutions which can also be motivated by subject-matter considerations: (a) The association effects could be omitted signalling that all teams are sufficiently different. Note that the full team effect then still captures association differences. (b) Alternatively, the team effect could be specified as a random effect (with zero mean) conveying that the winning logits for each team deviate randomly from the mean as captured by the remaining effects (e.g., by fixed association differences). (c) A random effect for the bookmakers would be conceivable implying that the bookmakers' odds deviate randomly from the mean as captured by the remaining effects. (d) Finally, the four different specifications of the deviation $\epsilon_{i,b}$ of bookmaker b for team i represent different views on the sources of variation and thus disagreement. For example, even if there is a fixed team effect α_i in the consensus, it would be conceivable that the amount of disagreement is only driven by the team's association because bookmakers might have a similar degree of information about teams in the same association. Combinations of the ideas (a)–(d) lead to 20 different mixed-effects models. Table 1 specifies the different effects and standard deviations of $\epsilon_{i,b}$ for each model. To find a parsimonious model which still gives a good approximation of the underlying data-generating process, standard model selection methods can be employed. We use the Bayesian Information Criterion (BIC).

3.2 Model selection

Fitting the 20 conceivable mixed-effects models discussed in the previous sections yields the results in Table 1 which provides the log-likelihood, number of parameters, and associated BIC. In general, the model selection approach shows that all models including fixed team effects perform clearly better than models with a random team effect, even if an additional association effect is included. Furthermore, the models with constant standard deviation are worse than all models using other standard deviation specifications. The best model emerging from the BIC selection is Model 3 (BIC = 83.47), containing only a fixed team effect (and hence no additional association) and a team-specific standard deviation. The second best model (Model 7) includes an additional random effect for the bookmakers, capturing bookmaker differences. The best four models (Models 3, 4, 7, and 8) have a fixed team effect and a team- or association-specific standard deviation. In summary, this conveys that, as expected, the main differences are across individual teams which require a full fixed effect (and can not be sufficiently captured by more parsimonious parametrizations such as a fixed association effect plus a random team effect). Furthermore, the fact that the bookmaker

Table 1: Effect and standard deviation specifications of the mixed-effects models for $\text{logit}(p_{i,b})$ of team i by bookmaker b. Each model is evaluated by the log-likelihood value (logLik), the number of estimated parameters (df), and the BIC.

	Team	Bookmaker	Association	Deviation	logLik	df	BIC
1	fixed	fixed	none	const	−2.51	63	439.70
2	fixed	none	none	const	−118.87	33	465.42
3	fixed	none	none	team	179.06	64	83.47
4	fixed	none	none	association	121.71	49	94.66
5	fixed	random	none	const	−50.86	34	336.32
6	fixed	random	none	bookmaker	12.72	64	416.14
7	fixed	random	none	team	179.01	65	90.46
8	fixed	random	none	association	121.92	50	101.15
9	random	fixed	none	const	−130.51	33	488.72
10	random	fixed	fixed	const	−95.98	49	530.04
11	random	fixed	none	bookmaker	−70.11	63	574.90
12	random	fixed	fixed	bookmaker	−35.71	79	616.50
13	random	fixed	none	team	59.04	64	323.51
14	random	fixed	fixed	team	93.48	80	365.02
15	random	fixed	none	association	13.18	49	311.73
16	random	fixed	fixed	association	47.64	65	353.21
17	random	none	none	const	−243.12	3	506.93
18	random	none	none	bookmaker	−163.14	33	553.97
19	random	none	none	team	45.19	34	144.20
20	random	none	none	association	−10.31	19	151.72
21	fixed	none	none	linear	84.33	34	65.92
22	fixed	none	none	power	112.05	35	17.39

effect can be omitted or captured as a random effect suggests that there are no large systematic deviations between the bookmakers. Similarly, a team-specific standard deviation is necessary to obtain the best model fit. However, models including association-specific standard deviations are only slightly worse, implying that agreement across bookmakers is driven to a large extent by the association differences.

Model 3 confirms the simple strategy of employing team-specific means for the consensus and team-specific standard deviations for agreement across bookmakers. It is reassuring that this intuitive model has been selected from a more general class of models, where some of the alternatives would have also had appealing interpretations. In Section 4.2 it is shown how the parametrization of the standard deviation can be made more parsimonious while retaining the same consensus (Models 21 and 22 of Table 1).

4 Analysis of the UEFA Champions League 2008/09

4.1 Consensus

The bookmaker consensus for the UEFA Champions League 2008/09 can be derived from the best model, Model 3 by using the estimated winning logits $\text{logit}(\widehat{p}_i) = \widehat{\nu} + \widehat{\alpha}_i$ which equal the team-specific means of the winning logits across the bookmakers for each team ($= 1/31 \sum_{b=1}^{31} \text{logit}(p_{i,b})$). This consensus information on the logit scale can be easily transformed to the associated winning probabilities $\widehat{p}_i$ of winning the tournament for all 32 participating teams which are shown in

Table 2: Estimated winning probabilities $\widehat{p}_i$, associated winning logits $\text{logit}(\widehat{p}_i)$ (reflecting the bookmakers consensus), and standard deviations $\widehat{\sigma}_i$ (reflecting the agreement across the bookmakers) for all 32 participating teams of the UEFA Champions League 2008/09. Additionally, the eight origin groups of the preliminaries, and the football association of the teams are shown.

	$\widehat{p}_i(\%)$	$\text{logit}(\widehat{p}_i)$	$\widehat{\sigma}_i$	Group	Association
Chelsea FC	13.71	−1.84	0.091	A	England
Manchester United FC	12.13	−1.98	0.092	E	England
FC Internationale Milano	10.16	−2.18	0.073	B	Italy
FC Barcelona	10.11	−2.19	0.066	C	Spain
Real Madrid CF	9.44	−2.26	0.162	H	Spain
Arsenal FC	6.39	−2.68	0.114	G	England
Liverpool FC	5.83	−2.78	0.111	D	England
FC Bayern München	4.58	−3.04	0.120	F	Germany
Juventus	3.84	−3.22	0.109	H	Italy
AS Roma	3.28	−3.38	0.086	A	Italy
FC Zenit St. Petersburg	2.48	−3.67	0.216	H	Russia
Olympique Lyonnais	2.46	−3.68	0.108	F	France
Club Atletico de Madrid	2.17	−3.81	0.176	D	Spain
Villarreal CF	1.93	−3.93	0.157	E	Spain
ACF Fiorentina	1.47	−4.20	0.171	F	Italy
Werder Bremen	1.30	−4.33	0.244	B	Germany
FC Porto	1.18	−4.43	0.317	G	Portugal
Olympique Marseille	0.81	−4.81	0.285	D	France
Fenerbahce SK	0.75	−4.89	0.149	G	Turkey
PSV Eindhoven	0.67	−4.99	0.309	D	Netherlands
FC Girondins de Bordeaux	0.61	−5.09	0.383	A	France
FC Shakhtar Donetsk	0.60	−5.11	0.330	C	Ukraine
Sporting Clube de Portugal	0.57	−5.16	0.319	C	Portugal
Panathinaikos FC	0.56	−5.17	0.260	B	Greece
FC Dynamo Kyiv	0.49	−5.31	0.435	G	Ukraine
Celtic FC	0.48	−5.33	0.217	E	Scotland
FC Steaua Bucuresti	0.32	−5.75	0.455	F	Romania
FC Basel 1893	0.21	−6.16	0.415	C	Switzerland
CFR 1907 Cluj	0.20	−6.20	0.453	A	Romania
Aalborg BK	0.14	−6.58	0.493	E	Denmark
Anaethosis Famagusta FC	0.11	−6.84	0.336	B	Cyprus
FC BATE Borisov	0.10	−6.89	0.405	H	Belarus

Table 2. Additionally, the eight origin groups of the preliminaries, and the football association of the teams are shown.

Chelsea FC is seen as the best team of the 32 teams and has the highest probability (13.71%) of winning the tournament. The expected runner-up (if the tournament schedule allows such a final) comes also from England, Manchester United FC (winning probability: 12.13%). The top two are followed by the champion of the "Serie A" FC Internationale Milano (10.16%) and the champion of the "Primera Division" FC Barcelona (10.11%). The last four teams are participating for the first time in the tournament and have just a winning probability of 0.20% or less. Using the group information in combination with the winning probabilities of the participating teams (Table 2) the following 16 teams (eight group-winners and eight runners-up) are expected to play

Table 3: Spearman's rank correlation between the actual tournament ranking, the ranking of the bookmaker consensus, the UEFA's seeding and the UEFA's club coefficient of the 32 participating teams.

	Bookmaker	Seeding	Coefficient
Tournament ranking	0.798	0.756	0.754
Bookmaker		0.843	0.841
Seeding			0.996

the first knock-out round: Chelsea FC, AS Roma (group A), FC Internationale Milano, Werder Bremen (B), FC Barcelona, FC Shakhtar Donetsk (C), Liverpool FC, Club Atletico de Madrid (D), Manchester United FC, Villarreal GF (E), FC Bayern München, Olympique Lyonnais (F), Arsenal FC, FC Porto (G), Real Madrid CF, and Juventus (H). In summary, the bookmaker consensus gives winning probabilities of the teams which can be used to predict the winner of the tournament. See Leitner *et al.* (2008) on how this forecast can be complemented for dynamics of such tournaments by a simulation approach.

To show how well the bookmaker consensus performs in practice, we compare the forecast with the real outcome of the UEFA Champions League 2008/09. Table 3 assesses the predictive performance of the bookmaker consensus by comparing them with the actual tournament outcome using Spearman's rank correlation. For the actual results, a total ranking including ties is employed, as commonly used in rankings of such incomplete tournaments. Various strategies for resolving the ties have been explored but did not lead to qualitatively different results. In addition, Table 3 also provides correlations with the ranking implied by the UEFA's seeding and UEFA's club coefficient of the teams (prior to the group drawn).

This shows that the bookmakers consensus has a very high correlation with the actual outcome (0.798) and performs somewhat better than the rankings based on the UEFA's seeding (0.756) and UEFA's club coefficient (0.754) of the teams. In particular, the bookmaker consensus correctly predicts three of four seminalists: Chelsea FC, Manchester United, FC Barcelona and 14 of 16 teams which played the first knockout round.

4.2 Agreement

In addition to the consensus of the bookmaker we can also derive the team-specific standard deviations of Model 3. As discussed above, the estimated standard deviations $\widehat{\sigma}_i$ captures the disagreement across the bookmakers. A low standard deviation for a team reflects a low disagreement across the bookmakers, whereas a high standard deviation implies a high disagreement across the bookmakers. The standard deviations σ_i for team i for all 32 participating teams are shown in Table 2.

In general, the team-specific standard deviations are low implying a low disagreement across the 31 bookmakers. The lowest disagreement across the bookmakers is with a standard deviation of 0.066 on the logit scale assigned to a top team, FC Barcelona. Conversely, the highest disagreement is with a standard deviation of 0.493 across bookmakers assigned to a team with a low consensus winning probability (Aalborg BK). Taking a closer look, we can see that the agreement increases with increasing winning logits of the teams brought out graphically in Figure 1. By exploiting this relationship information our current best model (Model 3) can be extended to a more parsimonious model modeling a relationship between the team-specific standard deviation and the fitted values on the logit scale:

$$\sigma_{i,b} = \sigma_i = \gamma_1 + \gamma_2 \text{logit}\,(p_i)^{\gamma_3}\,, \tag{5}$$

where γ_1, γ_2, and γ_3 are the function parameters which are estimated by the model.

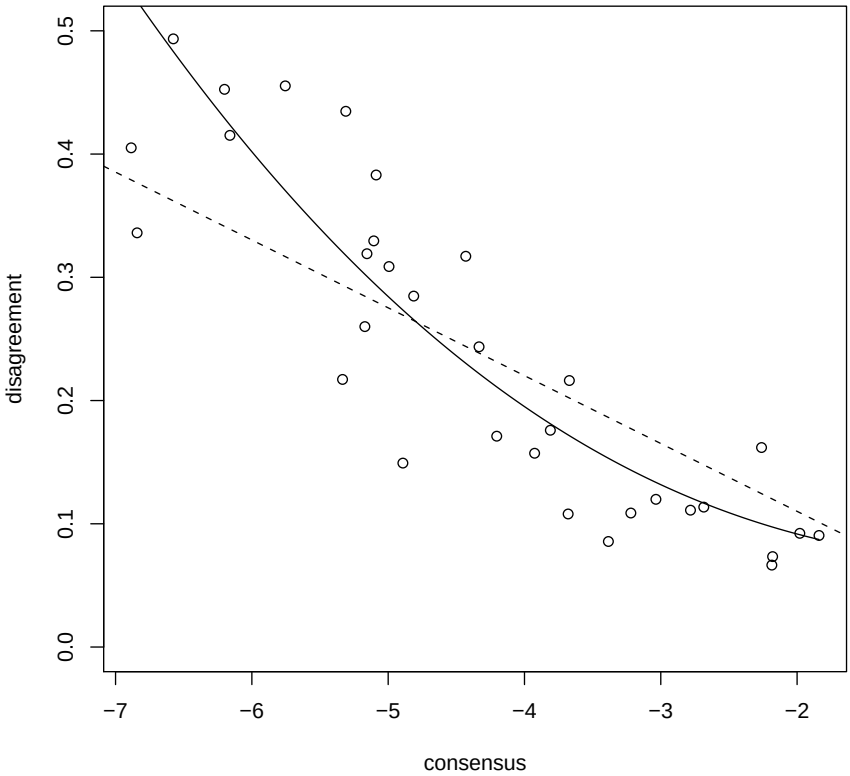

Figure 1: Relationship between the estimated bookmaker consensus (in winning logits) and agreement (standard deviation on the logit scale) of all 32 participating teams of the UEFA Champions League 2008/09. The points show the team-specific, the dashed line the linear and the solid line the non-linear relationship captured by the Models 3, 21 and 22 of Table 1.

In addition to the power specification above we also investigate a linear specification ($\gamma_3 = 1$). By using a linear relationship a much more parsimonious model, reducing the number of estimated parameters from 64 $(32 + 32)$ to 34 $(32 + 2)$ and improving the model selection criteria (BIC = 65.92) can be fitted (see Model 21 of Table 1). The estimated function parameters of the linear relationship are: $\gamma_1 = 0.000$ and $\gamma_2 = 0.055$. By estimating one more model parameter for the power γ_3 of a non-linear relationship the model can be improved again yielding a BIC of 17.39 (see Model 22 of Table 1). The estimated function parameters of the non-linear relationship are: $\gamma_1 = 0.067$, $\gamma_2 = 0.005$, and $\gamma_3 = 2.369$. Figure 1 shows the team-specific relationship of Model 3 (points), as well as the linear relationship of Model 21 (dashed line) and the non-linear relationship of Model 22 (solid line). Note that in all three models (Models 3, 21, and 22) all parameters are estimated simultaneous yielding the same estimated bookmaker consensus, but different specifications of disagreement across the bookmakers.

4.3 Team's association

According to the bookmaker consensus (Table 2) four teams out of the first seven ranked teams are from England which implies that England is the strongest European association. But what about

Table 4: Number of qualified teams, average consensus (in winning logits) and average disagreement (average standard deviation) for the 17 associations of all 32 participating teams of the UEFA Champions League 2008/09.

	No. of teams	Av. consensus	Av. disagreement
England	4	−2.32	0.102
Spain	4	−3.05	0.140
Italy	4	−3.25	0.110
Russia	1	−3.67	0.216
Germany	2	−3.68	0.182
France	3	−4.53	0.259
Portugal	2	−4.79	0.318
Turkey	1	−4.89	0.149
Netherlands	1	−4.99	0.309
Greece	1	−5.17	0.260
Ukraine	2	−5.21	0.382
Scotland	1	−5.33	0.217
Romania	2	−5.98	0.454
Switzerland	1	−6.16	0.415
Denmark	1	−6.58	0.493
Cyprus	1	−6.84	0.336
Belarus	1	−6.89	0.405

the other associations? The estimated consensus can also be used to rank the 17 associations of the participating teams. Therefore, we compute the means of the winning logits $\text{logit}(\widehat{p}_i)$ of all teams stemming from an association a (see Table 4). The difference of these means and the overall mean ν of all 32 participating teams can be seen as an implied "association effect" on the logit scale. In addition to the average consensus of an association, Table 4 shows the average disagreement (average standard deviations) and the number of qualified teams of the 17 associations.

There is a strong correlation between the average consensus on the logit scale and the number of qualified teams (0.75) implying that strong associations according to the bookmakers consensus have a higher number of qualified teams (cf., UEFA's determination strategy for the number of eligible teams in Union of European Football Associations, 2009). England, Spain and Italy have the maximum number of qualified teams (four), but England with the highest average consensus on the logit scale (−2.32) is the strongest European association. Russia with only one team (FC Zenit St. Petersburg) is rated better than Germany (two teams), France (three teams) and Portugal (two teams). The association with the weakest (average) consensus is clearly Belarus where the team with the lowest probability of winning the Champions League (FC BATE Borisov) comes from.

In addition to the relationship between the association effects and the number of qualified teams, we can also show the relationship between the agreement of the teams and their associations. Table 4 shows that the disagreement across the 31 bookmakers is very low for the teams stemming from the top three associations (England, Spain and Italy) and increases with the increasing average consensus.

5 Conclusion

This paper investigates a general model class for the unknown "true" logits for winning the UEFA Champions League 2008/09 based on quoted bookmakers odds stemming from 31 international bookmakers, reflecting the bookmakers consensus as forecasting tool and the agreement across the bookmakers. A linear mixed-effects model framework capturing different effects for the teams, the bookmakers as well as for the team's associations and allowing different specifications for the standard deviation leads to a variety of models. According to a models selection approach using the BIC the natural strategy of using the means of the winning logits as consensus and the team-specific standard deviation as measure for disagreement is appropriate. The estimated winning probabilities derived from the bookmaker consensus predicts the actual outcome very well (correlation of 0.798), somewhat better than UEFA's expectations (UEFA's seeding and UEFA's club coefficient). In particular, the bookmaker consensus model correctly predict three of four seminalists: Chelsea FC, Manchester United FC, FC Barcelona and 14 of 16 teams which played the first knockout round. Furthermore, the analysis of the bookmakers agreement implies a negative relationship between the estimated winning probabilities of a team and the disagreement across the bookmakers which can be modeled by a linear relationship or a non-linear relationship. Both extended models capturing these relationships reduce the number of estimated parameters of the model substantially and improve the model selection criteria. By analyzing the team's associations, we show a strong positive relationship between the number of teams stemming from an association and the average consensus of an association reflecting the UEFA's strategy of assigning more eligible teams to "stronger" associations and a strong negative relationship between the disagreement across the bookmakers and the average consensus of an association.

Computational details

All computations were carried out in the R system (version 2.8.1) for statistical computing (R Development Core Team, 2009). In particular, the R package nlme version 3.1-90 (Pinheiro *et al.*, 2008) was used for the estimation of the mixed-effects models (see Pinheiro and Bates, 2000).

References

Boulier BL, Stekler HO (2003). "Predicting the Outcomes of National Football League Games." *International Journal of Forecasting*, **19**, 257–270.

Clements MP (2008). "Consensus and Uncertainty: Using Forecast Probabilities of Output Declines." *International Journal of Forecasting*, **24**, 76–86.

Dixon M, Pope P (2004). "The Value of Statistical Forecasts in the UK Association Football Betting Market." *International Journal of Forecasting*, **20**, 697–711.

Forrest D, Goddard J, Simmons R (2005). "Odds-Setters as Forecasters: The Case of English Football." *International Journal of Forecasting*, **21**, 551–564.

Henery RJ (1999). "Measures of Over-Round in Performance Index Betting." *Journal of the Royal Statistical Society D*, **48**(3), 435–439.

Kolb R, Stekler HO (1996). "Is There a Consensus Among Financial Forecasters." *International Journal of Forecasting*, **12**, 455–464.

Lahiri K, Teigland C (1987). "On the Normality of Probability Distributions of Inflation and GNP Forecasts." *International Journal of Forecasting*, **3**, 269–279.

Leitner C, Zeileis A, Hornik K (2008). "Predicting the Winner of the EURO 2008 (A Statistical Investigation of Bookmakers Odds)." *Report 76*, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series. URL http://epub.wu-wien.ac.at/.

Leitner C, Zeileis A, Hornik K (2009). "Forecasting the Winner of the UEFA Champions League 2008/09." In R Koning, P Scarf (eds.), "Proceedings of the 2nd International Conference on Mathematics in Sport – IMA Sport 2009," pp. 94–99. ISBN: 979-0-905091-21-1.

Pinheiro J, Bates D (2000). *Mixed-Effects Models in S and S-PLUS*. Statistics and Comupting. Springer-Verlag New York, New York, USA. ISBN 0-387-98957-9.

Pinheiro J, Bates D, DebRoy S, Sarkar D (2008). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-92, URL http://CRAN.R-project.org/package=nlme.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Schnader M, Stekler HO (1991). "Do Consensus Forecasts Exist?" *International Journal of Forecasting*, **7**, 165–170.

Song C, Boulier BL, Stekler HO (2007). "The Comparative Accuracy of Judgmental and Model Forecasts of American Football Games." *International Journal of Forecasting*, **23**, 405–413.

Song C, Boulier BL, Stekler HO (2009). "Measuring Consensus in Binary Forecasts: NFL Game Predictions." *International Journal of Forecasting*, **25**, 182–191.

Spann M, Skiera B (2009). "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediciton Markets, Betting Odds and Tipsters." *Journal of Forecasting*, **28**, 55–72.

Su V, Su J (1975). "An Evaluation of ASA/NBER Business Outlook Survey Forecasts." *Explorations in Economic Research*, **2**, 588–618.

Union of European Football Associations (2009). "UEFA Champions League." [Online; accessed 2009-03-23], URL http://www.uefa.com/competitions/ucl/.

Zarnowitz V, Lambros LA (1987). "Consensus and Uncertainty in Economic Prediction." *Journal of Political Economy*, **95**, 561–621.