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

**Christoph Leitner**
WU Wirtschaftsuniversität Wien

**Achim Zeileis**
Universität Innsbruck

**Kurt Hornik**
WU Wirtschaftsuniversität Wien

**Abstract**

Bookmakers odds are an easily available source of "prospective" information that is often employed for forecasting the outcome of sports events. In order 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 of 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 agreement across the bookmakers 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 take a somewhat different approach and employ statistical models to explore the heterogeneity in bookmakers' expectations as reflected in their quoted odds. The idea is to capture effects in means and variances of these expectations that can be related naturally to "consensus" and "disagreement" among the bookmakers. The resulting model predictions for the means can then canonically be employed as consensus forecasts and thus relate our work to (1) in the sense above. The approach is illustrated in an analysis of quoted long-term odds for winning the UEFA Champions League 2008/09 for all 32 participating teams by 31 international bookmakers.

In sports betting, bookmakers odds are prospective ratings of the performance of the participating players or teams in a sports competition which vary between the bookmakers. They have been successfully used to predict the outcome of single games (e.g., Spann and Skiera 2009; Song, Boulier, and Stekler 2007; Forrest, Goddard, and Simmons 2005; Dixon and Pope 2004; Boulier and Stekler 2003). Based on these ideas, Leitner, Zeileis, and Hornik (2009a,b) use aggregated quoted odds of a variety of bookmakers to forecast the outcome of whole tour-

This is a preprint of an article published in *IMA Journal of Management Mathematics*, **22**(2), 183–194. doi:10.1093/imaman/dpq016
Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

naments, 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.

Various strategies for aggregating information from different forecasters have been proposed in the literature. 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. Consensus forecasts can be computed as the median (Su and Su 1975) or the mean of all the forecasts in the sample (Zarnowitz and Lambros 1987). The latter is successfully applied by Leitner *et al.* (2009a); Leitner, Zeileis, and Hornik (2009c) to sports competitions. Alternative strategies for the aggregation of forecasts are discussed by Kolb and Stekler (1996) and Schnader and Stekler (1991). In order to measure “uncertainty” or “disagreement”, the standard deviations of the predictive probability distributions are typically used (e.g, Clements 2008; Zarnowitz and Lambros 1987; Lahiri and Teigland 1987). Furthermore, for the case of inter-rater agreement involving binary choices, Song, Boulier, and Stekler (2009) employ Cohen’s kappa coefficient to evaluate forecasts of National Football League games.

Here, we follow Leitner *et al.* (2009b) and extend their framework for modeling bookmakers odds to a more general model class. The models are based on the bookmakers’ expected winning probabilities derived from the raw quoted odds. As these probabilities are necessarily in the unit interval, straightforward linear modeling is not appropriate. We follow the standard technique of employing a suitable link function to transform probabilities to the real line and then using standard linear regressions or rather linear mixed-effects models with normally distributed errors as a generalization thereof. This naturally yields consensus forecasts and (dis)agreement measures as means and variances on the transformed scale, thus providing a convenient statistical framework for the aggregation of bookmakers odds.

Based on bookmakers odds for the occurence of a set of events (e.g., players/teams winning a particular match/tournament), a natural strategy for the computation of consensus and (dis)agreement are event-specific means and variances across the different bookmakers. The statistical modeling framework outlined above contains this strategy as a special case – namely fixed event effects for both means and variances – but also allows exploration of a wider range of model specifications. For example, potential advantages of random vs. fixed effects can be investigated, or effects pertaining to the bookmaker, grouping effects for the different events, or associations between means and variances can be exploited to specify more parsimonious models. In the application to the UEFA Champions League 2008/09, it can be shown that the straightforward strategy of event-specific means and variances performs well in a wide range of models. However it can be improved even further when the association between means and variances is incorporated, i.e., when considering that events with higher probability of occurence also have a higher level of agreement. The resulting bookmaker consensus forecast for the UEFA Champions League 2008/09 performs well in practice, exhibiting a high correlation with the actual tournament outcome.

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.

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

# 2. Tournament and data description

## 2.1. Tournament

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. Every year, a selection of European football clubs compete in a multi-stage format (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's 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 (quoted as decimal odds) for 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 out of 50 European top-selling online sports bookmakers who offer odds for this event. Figure 1 shows the quoted odds (on a log-axis) for all 32 participating teams of the UEFA Champions League 2008/09 by the 31 bookmakers. It can be seen that the heterogeneity increases along with the level of the quotes odds.

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). Assuming that each bookmaker $b = 1, \dots, 31$ applies the same overround $\delta_b$ for every team, the implied expected winning probabilities $p_{i,b}$ for team $i = 1, \dots, 32$ by bookmaker $b$ can be obtained from the raw quoted odds $rawodds_{i,b}$ via

$$p_{i,b} = \frac{1}{rawodds_{i,b}\,(1 + \delta_b)}, \tag{2.1}$$

where $\delta_b$ is chosen such that $\sum_i p_{i,b} = 1$. For our dataset we obtain a mean overround of 23.58% across all bookmakers with an interquartile range from 19.71% to 26.89%.

*UEFA's club coefficient and seeding*

The UEFA also announces their expectancies for the tournament outcome prior to the tournament by publishing a group draw seeding which is a ranking that is very similar to the

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

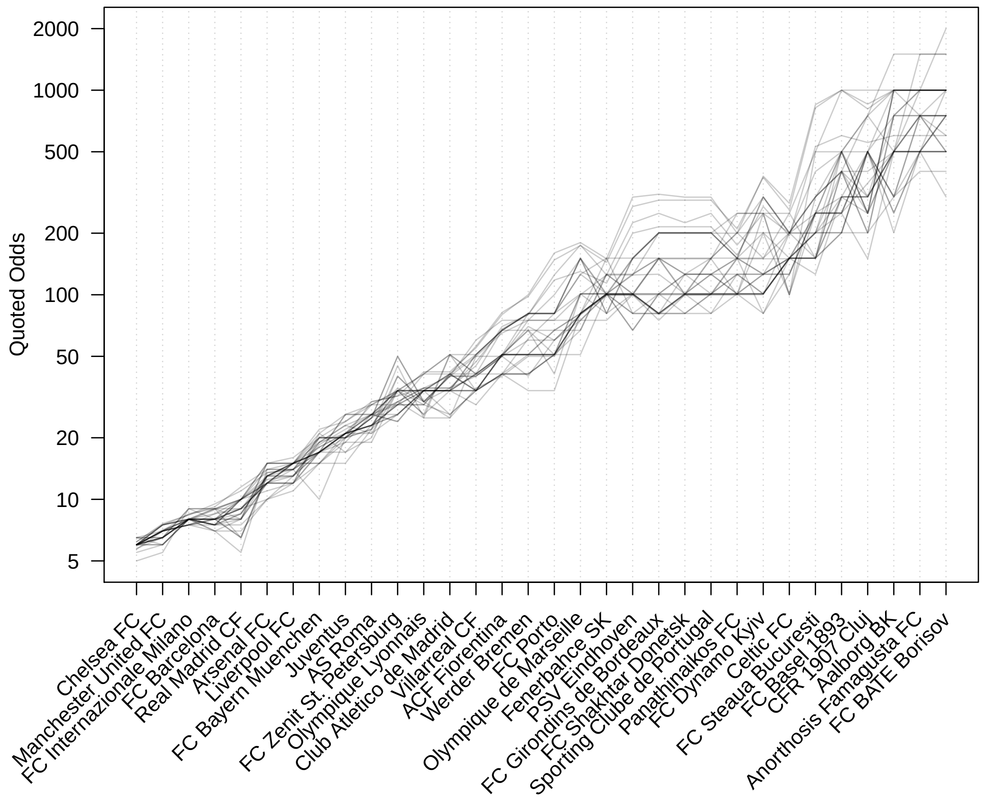


Figure 1: Quoted odds (on log-axis) for all 32 participating teams of the UEFA Champions League 2008/09 by the 31 bookmakers.

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 competitions in the last five seasons, and the league coefficient. The latter is also used to determine the number of eligible teams for the UEFA Champions League where the best three associations have four teams in the tournament (for more details see Union of European Football Associations 2009). We obtained the UEFA's club coefficient and seeding for the group draw on 2008-08-28 from UEFA's website for all 32 participating teams and, in Section 4, compare both to the ranking derived from the bookmakers' consensus forecast.

## 3. Modeling consensus and agreement

### 3.1. Model class

To model the expected winning probabilities $p_{i,b}$ for each team $i = 1, \dots, 32$ and bookmaker $b = 1, \dots, 31$, as derived from the raw quoted odds, straightforward linear models are not appropriate as the $p_{i,b}$ necessarily lie within the unit interval. Therefore, we follow the standard

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

technique of employing a suitable link function to transform probabilities to the real line and then using linear models for the transformed data. Various link functions are conceivable; standard choices include the logit or probit link function. In the following, we employ the logit link throughout; using the probit link instead would lead to qualitatively similar results.

On the transformed logit scale, an intuitive and straightfoward strategy would be to compute team-wise means for the consensus and team-wise standard deviations for the disagreement across bookmakers (as suggested by, e.g., Zarnowitz and Lambros 1987). In our application, this simple strategy might be appropriate because we could expect the teams to be sufficiently different and the bookmakers to have rather similar information about the teams. However, from a statistical point of view one should investigate whether 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 still giving a good approximation of the underlying data-generating process. Therefore, we propose a stochastic model class that captures the underlying probability distribution on a logit scale and contains the simple strategy as a special case. We assume additive and normally distributed "errors" on the logit scale, providing a natural way for assessment of means and variances in the models.

The model relates the expected 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) normally-distributed error term $\epsilon_{i,b}$ of bookmaker $b$ for team $i$, reflecting the disagreement across the bookmakers. In order to capture these latent quantities by a linear mixed-effects model, we allow the true winning logits to depend on a team effect $\alpha_i$, an association effect $\beta_{a(i)}$ for association $a$ of team $i$, as well as an overall intercept $\nu$. The error can additionally depend on $\mu_b$, the mean effect of bookmaker $b$. We also allow different specifications of the standard deviation $\sigma_{i,b}$ of bookmaker $b$ for team $i$. In summary, this can be written as

$$\begin{aligned} \text{logit}(p_{i,b}) &= \text{logit}(p_i) + \epsilon_{i,b} && (3.1) \\ &= \nu + \alpha_i + \beta_{a(i)} + \mu_b + \sigma_{i,b} Z_{i,b}, && (3.2) \end{aligned}$$

where $Z_{i,b}$ is a standardized error and $\sigma_{i,b}$ is the standard deviation which can either be 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 $\alpha_i$, $\beta_{a(i)}$, and $\mu_b$ are included as fixed effects due to the dependence of association $a(i)$ on the team $i$.

In order 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 $\alpha_i$ 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

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

Table 1: Effect and standard deviation specifications of the mixed-effects models for 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 $\alpha_i$ | Bookmaker $\mu_b$ | Association $\beta_{a(i)}$ | Deviation $\sigma_{i,b}$ | logLik | df | BIC |
|---|---|---|---|---|---|---|---|
| 1 | fixed | fixed | none | const | −3.20 | 63 | 441.09 |
| 2 | fixed | none | none | const | −121.71 | 33 | 471.11 |
| 3 | fixed | none | none | team | 179.73 | 64 | 82.13 |
| 4 | fixed | none | none | association | 121.48 | 49 | 95.12 |
| 5 | fixed | random | none | const | −51.88 | 34 | 338.34 |
| 6 | fixed | random | none | bookmaker | 12.61 | 64 | 416.37 |
| 7 | fixed | random | none | team | 179.73 | 65 | 89.03 |
| 8 | fixed | random | none | association | 121.63 | 50 | 101.72 |
| 9 | random | fixed | none | const | −130.99 | 33 | 489.68 |
| 10 | random | fixed | fixed | const | −96.30 | 49 | 530.69 |
| 11 | random | fixed | none | bookmaker | −69.91 | 63 | 574.51 |
| 12 | random | fixed | fixed | bookmaker | −35.35 | 79 | 615.78 |
| 13 | random | fixed | none | team | 59.08 | 64 | 323.41 |
| 14 | random | fixed | fixed | team | 93.68 | 80 | 364.62 |
| 15 | random | fixed | none | association | 12.88 | 49 | 312.32 |
| 16 | random | fixed | fixed | association | 47.49 | 65 | 353.50 |
| 17 | random | none | none | const | −245.68 | 3 | 512.05 |
| 18 | random | none | none | bookmaker | −163.39 | 33 | 554.47 |
| 19 | random | none | none | team | 46.04 | 34 | 142.51 |
| 20 | random | none | none | association | −10.33 | 19 | 151.75 |
| 21 | fixed | none | none | linear | 83.35 | 34 | 67.88 |
| 22 | fixed | none | none | power | 113.47 | 35 | 14.56 |

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. In order 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; Pinheiro and Bates 2000).

## 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. With respect to the BIC, the best model emerging from Models 1–20 is Model 3 (BIC = 82.13), 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

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

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 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 easily be transformed to the associated winning probabilities $\widehat{p}_i$ of winning the tournament for all 32 participating teams which are shown in 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 (13.52%) 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: 12.00%). The top two are followed by FC Internazionale Milano (10.10%) and FC Barcelona (10.05%). The last four teams are participating for the first time in the tournament and have just a winning probability of 0.21% 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 the first knock-out round: Chelsea FC, AS Roma (group A), FC Internazionale Milano, Werder Bremen (B), FC Barcelona, FC Shakhtar Donetsk (C), Liverpool FC, Club Atlético 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.* (2009a) on how this forecast can be complemented for dynamics of such tournaments by a simulation approach.

In order 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

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

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 | 13.52 | −1.86 | 0.092 | A | England |
| Manchester United FC | 12.00 | −1.99 | 0.091 | E | England |
| FC Internazionale Milano | 10.10 | −2.19 | 0.074 | B | Italy |
| FC Barcelona | 10.05 | −2.19 | 0.065 | C | Spain |
| Real Madrid CF | 9.40 | −2.27 | 0.157 | H | Spain |
| Arsenal FC | 6.41 | −2.68 | 0.111 | G | England |
| Liverpool FC | 5.86 | −2.78 | 0.109 | D | England |
| FC Bayern München | 4.62 | −3.03 | 0.119 | F | Germany |
| Juventus | 3.88 | −3.21 | 0.107 | H | Italy |
| AS Roma | 3.32 | −3.37 | 0.085 | A | Italy |
| FC Zenit St. Petersburg | 2.52 | −3.66 | 0.213 | H | Russia |
| Olympique Lyonnais | 2.49 | −3.67 | 0.108 | F | France |
| Club Atlético de Madrid | 2.20 | −3.80 | 0.175 | D | Spain |
| Villarreal CF | 1.96 | −3.91 | 0.157 | E | Spain |
| ACF Fiorentina | 1.50 | −4.19 | 0.173 | F | Italy |
| Werder Bremen | 1.32 | −4.32 | 0.245 | B | Germany |
| FC Porto | 1.20 | −4.41 | 0.319 | G | Portugal |
| Olympique de Marseille | 0.82 | −4.79 | 0.286 | D | France |
| Fenerbahçe SK | 0.76 | −4.87 | 0.151 | G | Turkey |
| PSV Eindhoven | 0.69 | −4.98 | 0.312 | D | Netherlands |
| FC Girondins de Bordeaux | 0.62 | −5.07 | 0.385 | A | France |
| FC Shakhtar Donetsk | 0.61 | −5.09 | 0.333 | C | Ukraine |
| Sporting Clube de Portugal | 0.58 | −5.14 | 0.321 | C | Portugal |
| Panathinaikos FC | 0.58 | −5.15 | 0.261 | B | Greece |
| FC Dynamo Kyiv | 0.50 | −5.29 | 0.437 | G | Ukraine |
| Celtic FC | 0.49 | −5.32 | 0.221 | E | Scotland |
| FC Steaua Bucureşti | 0.32 | −5.74 | 0.457 | F | Romania |
| FC Basel 1893 | 0.21 | −6.14 | 0.417 | C | Switzerland |
| CFR 1907 Cluj | 0.21 | −6.18 | 0.456 | A | Romania |
| Aalborg BK | 0.14 | −6.56 | 0.494 | E | Denmark |
| Anorthosis Famagusta FC | 0.11 | −6.82 | 0.336 | B | Cyprus |
| FC BATE Borisov | 0.10 | −6.87 | 0.405 | H | Belarus |

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

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

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 |

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 semifinalists (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 $\sigma_i$ 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 team with the lowest disagreement across the bookmakers is one of the top teams, FC Barcelona, with a standard deviation of 0.065 on the logit scale. Conversely, the team with the highest disagreement (standard deviation 0.494) is Aalborg BK which has a low consensus winning probability. Taking a closer look (see Figure 2), we can see that the agreement increases with increasing winning logits of the teams. By exploiting this information, our current best model (Model 3) can be improved further by fitting a relationship between the team-specific standard deviations and the fitted values on the logit scale:

$$\sigma_{i,b} = \sigma_i = \gamma_1 + \gamma_2 \text{logit}\,(p_i)^{\gamma_3}\,, \tag{4.1}$$

where $\gamma_1$, $\gamma_2$, and $\gamma_3$ are the function parameters which are estimated by the model (jointly along with the parameters specifying the consensus $\text{logit}(p_i)$).

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 criterion (BIC = 67.88) 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 $\gamma_3$ of a non-linear relationship the model can be improved again yielding a BIC of 14.56 (see Model 22 of Table 1). The estimated function parameters of the non-linear

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

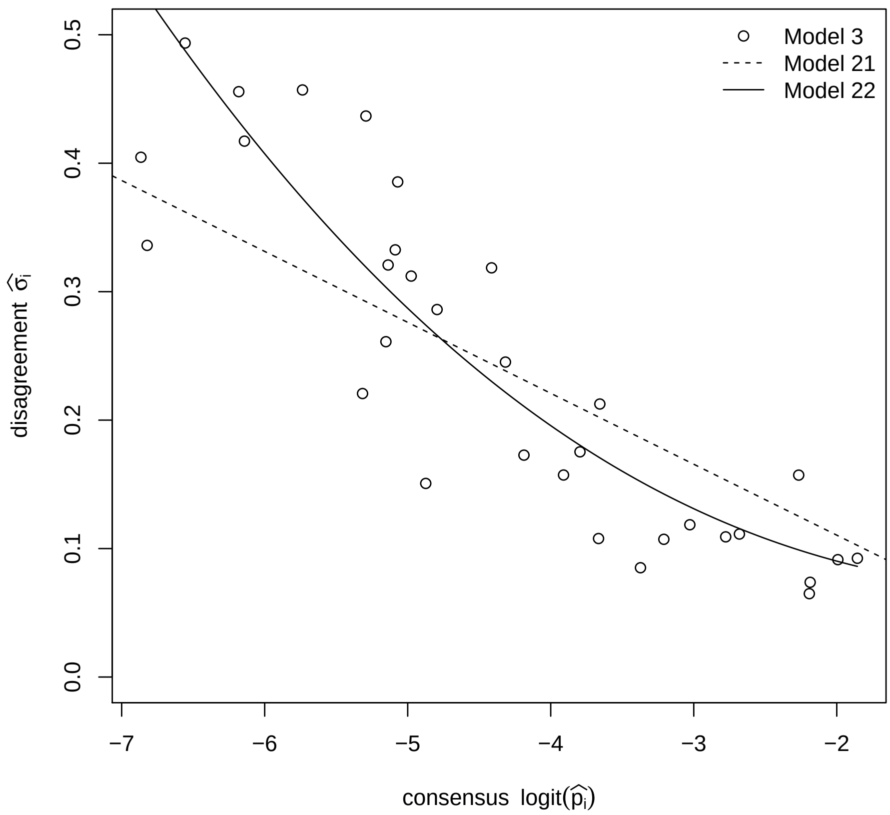


Figure 2: Relationship between the estimated bookmaker consensus logit($\widehat{p}_i$) and different specifications of disagreement $\widehat{\sigma}_i$ for all 32 participating teams of the UEFA Champions League 2008/09. The points show the team-specific, the dashed line the linear and the solid line the non-linear relationship captured by the Models 3, 21 and 22 of Table 1.

relationship are: $\gamma_1 = 0.065$, $\gamma_2 = 0.005$, and $\gamma_3 = 2.375$. Figure 2 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 simultaneously 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 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 logit($\widehat{p}_i$) of all teams coming from an association $a$ (see Table 4). The difference of these means and the overall mean $\nu$ of all 32 participating teams can be seen as an implied

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

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 | −2.33 | 0.101 |
| Spain | 4 | −3.04 | 0.139 |
| Italy | 4 | −3.24 | 0.110 |
| Russia | 1 | −3.66 | 0.213 |
| Germany | 2 | −3.67 | 0.182 |
| France | 3 | −4.51 | 0.260 |
| Portugal | 2 | −4.78 | 0.320 |
| Turkey | 1 | −4.87 | 0.151 |
| Netherlands | 1 | −4.98 | 0.312 |
| Greece | 1 | −5.15 | 0.261 |
| Ukraine | 2 | −5.19 | 0.385 |
| Scotland | 1 | −5.32 | 0.221 |
| Romania | 2 | −5.96 | 0.456 |
| Switzerland | 1 | −6.14 | 0.417 |
| Denmark | 1 | −6.56 | 0.494 |
| Cyprus | 1 | −6.82 | 0.336 |
| Belarus | 1 | −6.87 | 0.405 |

“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 (−2.33) 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 coming from the top three associations (England, Spain and Italy) and increases with the increasing average consensus.

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

## 5. Conclusion

Based on quoted bookmakers odds for the occurences of a certain set of events (such as players/teams winning a particular sports match/tournament), this paper investigates a general model class for the unknown "true" logits of the occurence of the events. It is applied to the assessment of consensus and (dis)agreement among 31 international bookmakers for the UEFA Champions League 2008/09. 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 deviations leads to a variety of models. According to a model 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 predict 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 predicts three of four semifinalists (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 criterion. By analyzing the team's associations, we show a strong positive relationship between the number of teams coming from an association and the average consensus of the respective association. This reflects UEFA's strategy of allocating more fixed and qualifying slots to "stronger" associations. Finally, we find 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 2.9.2) for statistical computing (R Development Core Team 2010). In particular, the R package nlme version 3.1-92 (Pinheiro, Bates, DebRoy, and Sarkar 2009) 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.

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

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 (2009a). "Forecasting Sports Tournaments by Ratings of (Prob)abilities: A Comparison for the EURO 2008." *International Journal of Forecasting*, **26**(3), 471–481. doi:10.1016/j.ijforecast.2009.10.001.

Leitner C, Zeileis A, Hornik K (2009b). "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.

Leitner C, Zeileis A, Hornik K (2009c). "Is Federer Stronger in a Tournament Without Nadal? An Evaluation of Odds and Seedings for Wimbledon 2009." *Austrian Journal of Statistics*, **38**(4), 277–286.

Pinheiro J, Bates D (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York, USA.

Pinheiro J, Bates D, DebRoy S, Sarkar D (2009). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-96, URL http://CRAN.R-project.org/package=nlme.

R Development Core Team (2010). *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 Prediction 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.

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

**Affiliation:**
Christoph Leitner, Kurt Hornik
Institute for Statistics and Mathematics
WU Wirtschaftsuniversität Wien
Augasse 2–6
1090 Wien, Austria
E-mail: Christoph.Leitner@wu.ac.at, Kurt.Hornik@wu.ac.at

Achim Zeileis
Department of Statistics
Universität Innsbruck
Universitätsstr. 15
6020 Innsbruck, Austria
E-mail: Achim.Zeileis@R-project.org

Copyright © 2010 The Authors. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.