Received: 27 January 2017 | Revised: 25 November 2017 | Accepted: 12 January 2018

DOI: 10.1111/stan.12127

ORIGINAL ARTICLE

WILEY

A Bayesian asymmetric logistic model of factors underlying team success in top-level basketball in Spain

José María Pérez-Sánchez[1] | Román Salmerón-Gómez[2] | Francisco M. Ocaña-Peinado[3]

[1]Campus Universitario de Tafira, s/n. Department of Applied Economic Analysis, University of Las Palmas de Gran Canaria, Las Palmas 35017, Spain

[2]Campus Universitario de Cartuja, s/n. Department of Quantitative Methods for Economics and Business, University of Granada, Granada, 18071, Spain

[3]Campus Universitario de Cartuja, s/n. Department of Statistics and Operation Research, University of Granada Granada 18071, Spain

Correspondence
José María Pérez-Sánchez, Campus Universitario de Tafira, s/n. Department of Applied Economic Analysis, University of Las Palmas de Gran Canaria, Las Palmas 35017, Spain.
Email: josemaria.perez@ulpgc.es

Funding information
Ministerio de Economía y Competitividad, Spain, Grant/Award Number: ECO2013–47092; Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación, Spain, Grant/Award Number: ECO2017–85577–P

This paper analyses the factors underlying the victories and defeats of the Spanish basketball teams Real Madrid and Barcelona in the national league, ACB. The following research questions were addressed: (a) Is it possible to identify the factors underlying these results? (b) Can knowledge of these factors increase the probability of winning and thus help coaches take better decisions? We analysed 80 and 79 games played in the 2013–2014 season by Real Madrid and Barcelona, respectively. Logistic regression analysis was performed to predict the probability of the team winning. The models were estimated by standard (frequentist) and Bayesian methods, taking into account the asymmetry of the data, that is, the fact that the database contained many more wins than losses. Thus, the analysis consisted of an asymmetric logistic regression. From the Bayesian standpoint, this model was considered the most appropriate, as it highlighted relevant factors that might remain undetected by standard logistic regression. The prediction quality of the models obtained was tested by application to the results produced in the following season (2014–2015). Again, asymmetric logistic regression achieved the best results. In view of the study findings, we make various practical recommendations to improve decision making in this field. In short, asymmetric logistic regression is a valuable tool that can help coaches improve their game strategies.

KEYWORDS
asymmetric link, Bayesian estimation, logistic regression, model selection

1 | INTRODUCTION

In sports clubs, as in all organisations, decisions must be taken in order to maintain and, if possible, improve performance. Quantitative methods can be used to assist in the decision-making process and thus avoid excessive subjectivity. In the context of professional sports, the emergence of the "Moneyball phenomenon" (Lewis, 2004) has produced many changes in the use of information sources. The main focus of this approach is to measure the contributions made by each player, from the data generated during a game, in order to optimise available resources and thus improve the team's performance.

Sports decisions, regarding the team design, transfers, or new signings (of players or coaches), are restricted by the economic limitations facing the organisation. Clearly, wealthier clubs have greater possibilities of signing elite athletes and thus of buying better performance. Clubs with fewer resources, however, might counter this imbalanced situation by analysing quantitative information in order to sign the best human resources possible in accordance with the available budget or to detect and exploit their opponents' weaknesses. However, the "Moneyball philosophy" cannot readily be extrapolated to basketball, due to the great complexity of this sport. Innumerable factors may influence the performance of a player and his team, such as the opponent (defence tactics, types of players, pace, etc.), teammates' abilities, the team's playing style, the coach's philosophy, and the stage of the match. Certainly, it is very complicated to measure all these aspects using only the traditional box score. In response, much has been done to improve the quality and quantity of available information, and a key reference study in this respect was presented by Oliver (2004). In the United States, changes in this field are exemplified by the National Basketball Association (NBA), which has adopted advanced systems to measure the performance of teams and players, considering many more elements than the traditional statistics referring to scoring, shooting percentages, and rebounds (see http://stats.nba.com/). Many studies have been conducted of basketball from a quantitative standpoint. Some of these are based on descriptive statistics, probability and inference (Bar–Eli & Tractinsky, 2000; Csapo, Avugos, Raab & Bar-Eli, 2015; Del Corral, Maroto & Gallardo, 2017; Gómez et al., 2015; Martínez & Martínez, 2010; Padulo et al., 2015a, 2015b; Stoszkowski & Collins, 2016; Viggiano et al., 2014), discriminant analysis (García, Ibáñez, Martínez De Santos, Leite & Sampaio, 2013; García, Ibáñez, Gómez & Sampaio, 2014; Gómez, Lorenzo, Sampaio & Ibáñez, 2006; Gómez, Lorenzo, Ortega, Sampaio & Ibáñez, 2009; Ibáñez et al., 2008; Lorenzo, Gómez, Ortega, Ibáñez & Sampaio, 2010; Sampaio & Janeira, 2003), multiple regression models (Casals & Martínez, 2013; Gómez & Salmerón, 2016; Martínez & Caudill, 2013; Mikolajec, Maszczyk & Zajac, 2013), binary regression models (Del Corral, García–Unanue & Herencia, 2016; Gómez, Lorenzo, Ibáñez & Sampaio, 2013; Sánchez, Castellanos & Dopicoa, 2007; Stekler & Klein, 2012), panel data models (De La Torre–Ruiz, & Aragón–Correa, 2012; Giambatista, 2004; Gómez & Salmerón, 2015), linear programming techniques (Fizel & D'itri, 1996; McGoldrick & Voeks, 2004), or structural equation modelling (Baghal, 2012).

The main goal of this paper is to contribute to the empirical literature in this field by addressing the following research questions. First, is it possible to identify the factors that determine whether a basketball team will win or lose? Second, can this information be used to increase the probability of winning? This study, based on an analysis of the victories and defeats of the Real Madrid and Barcelona teams in the Spanish ACB, is expected to be of special interest to basketball team managers, but our findings may also provide useful information to team managers in other sports and help improve their strategic decision making. In this context, Karipidis, Fotinakis, Taxildaris, and Fatouros (2001) analysed the basketball matches that took place during the 1997 and 1999 Eurobasket tournament, the 1996 Olympic Games and the 1998 World Basketball Championship. In the same regard, Sampaio, Janeira, Ibáñez, and Lorenzo (2006) examined the differences in game-related statistics between basketball players playing in professional leagues in the United States, Spain, and Portugal and conducted a performance analysis. According to Sampaio, Lago–Peñas, and Drinkwater (2008), whose study aim was to identify such factors, the success of the U.S.A. team in the 2008 Olympic Games was related to its speed of play. Sánchez Castellanos & Dopicoa (2007) derived a logistic model, using a database of Spanish basketball statistics to make an empirical evaluation of the relative importance of various factors in team performance. Ibáñez et al. (2008) later identified game-related statistics highlighting the differences, over the season, between successful and unsuccessful ACB teams, and concluded that the key factors were passing skills and defensive preparation. Sampaio, Drinkwater, and Leitea (2010) created individual performance profiles based on team quality and playing time (each of the four periods, or quarters, into which the playing time is divided) and concluded that the periods are not related to these performance profiles. De La Torre–Ruiz and Aragón–Correa (2012) focused on the performance of newcomers in NBA teams. Gómez Lorenzo, Ibáñez & Sampaio (2013) identified various performance indicators as predictors of the effectiveness of ball possession in men's and women's basketball, in an analysis of matches in the Spanish ACB. Mikolajec, Maszczyk, & Zajac (2013) attempted to identify the team factors associated with success in the NBA league. Teramoto and Cross (2010) used multiple linear regression and logistic regression analysis to determine which factors may increase the probability of winning games in the NBA, distinguishing between the first ("regular") phase of the season and the playoffs. Summers (2013) studied factors related to success in the NBA playoffs. Both of the latter studies, as well as those by Baghal (2012) and Kubatko, Oliver, Pelton, and Rosenbaum (2007), include in their analysis the four factors identified by Oliver (2004): shooting efficiency, number of turnovers, offensive rebounds, and free throws made.

However, to our knowledge, no previous publications have considered the situation in which one basketball team has many more wins than losses. A standard logit model can be used to analyse the factors that determine sporting performance, but sometimes, the individual results are clearly related more to one category than to another. This is the case in the present context, in which the database is unbalanced, containing many more wins than losses. In this situation, the use of an asymmetric or skewed logit model can improve the estimation results. In this respect, Chen, Dey, and Shao (1999) applied a Bayesian approach, using an asymmetric link in their analysis of binary response data when one response is much more frequent than the other. Similarly, Bermúdez, Pérez, Ayuso, Gómez, and Vázquez (2008) applied asymmetric logistic regression to model the fraudulent behaviour reflected in a database of insurance claims in Spain. In the area of health care, Sáez–Castillo et al. (2010) used an asymmetric logistic approach to analyse infection rates in a general and digestive surgery hospital department. More recently, Pérez–Sánchez, Negrín–Hernández, García–García, and Gómez–Déniz (2014) analysed the risk factors underlying automobile insurance claims by considering the skewed link function in a logistic regression model.

In our study, the data set presented nearmulticollinearity, a problem that was addressed by means of orthogonal regression (Novales, Salmerón, García, García & López, 2015). This technique not only alleviates multicollinearity but also allows the results to be interpreted from a different point of view. On the basis of this analysis, basketball team managers can obtain greater knowledge about the relevant factors and thus take better decisions.

The rest of this paper is structured as follows: Section 2 presents the theoretical aspects of the logistic and orthogonal regressions considered. The variables selected for the econometric models are derived from the analysis shown in Section 3. In the following section, the results obtained are described and compared. Finally, Section 5 summarises the main conclusions drawn and highlights the contributions of this paper.

2 | LOGISTIC REGRESSION

In this section, we propose two alternative logit models to fit the probability of a win. First, frequentist and Bayesian approaches are used to estimate a standard logistic model; then we assume an asymmetric link from the Bayesian point of view.

2.1 | Frequentist estimation of logit models

Let $y = (y_1, y_2, \ldots, y_n)'$ denote an $n \times 1$ vector of a dependent dichotomous variable, and let $x_i = (x_{i1}, \ldots, x_{ik})'$ denote the $k \times 1$ vector of covariates for the match i. The problem of estimating the probability of belonging to a group included in y_i is then addressed by fitting a regression model. In this study, if $y_i = 1$, the ith team wins a match, and $y_i = 0$ otherwise. We assume that $y_i = 1$ with probability p_i and $y_i = 0$ with probability $1 - p_i$. The regression model is given by $p_i = F(x_i'\beta)$, where F is the inverse of the standard logistic cumulative function (link function) and $\beta = (\beta_1, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients, which represents the effect of each variable x_i in the model. Thus, the likelihood is given by

$$l(y|x, \beta) = \prod_{i=1}^{n} [F(x_i'\beta)]^{y_i} [1 - F(x_i'\beta)]^{1-y_i}, \tag{1}$$

where $F(s) = 1/(1+e^{-s}), -\infty < s < \infty$ is a symmetric function with respect to zero. The regression coefficients are usually estimated by numerical evaluation of the likelihood function. In this way, the model provides the probability of each team winning a basketball game. The next step is to consider a cutoff in this probability in order to determine whether a team will win or not. The standard logistic model was evaluated using STATA econometric software (2015).

2.2 | Bayesian estimation of logit models

In this case, the regression coefficients are considered to be random variables. We propose two different models, both of which are considered using Bayesian estimators, that is, symmetric and asymmetric link functions are used, in both cases assuming noninformative and centred normal prior distributions for the β coefficients in order to facilitate comparison with frequentist estimations. The use of an asymmetric link function is recommended for binary response data when one response is much more frequent than the other, see Stukel (1988) and Chen, Dey and Shao (1999). From the asymmetric standpoint, an approach based on data augmentation, as considered by Albert and Chib (1993), can be used. In this way, it is easily shown that the skewed logit link is equivalent to considering that

$$y_i = \begin{cases} 1, & w_i \geq 0; \\ 0, & w_i < 0, \end{cases} \tag{2}$$

where $w_i = xi'\beta + \delta z_i + \epsilon_i, z_i \sim G, \epsilon_i \sim F$. We assume that z_i and ϵ_i are independent and that F is the standard logistic cumulative distribution function. Moreover, G is the cumulative distribution function of the half-standard normal distribution given by

$$g(z) = \frac{2}{\sqrt{2\pi}} e^{-z^2/2}, \quad z > 0.$$

In this model, $\delta \in (-\infty, \infty)$ is the skewness parameter. Thus, the skewness of the regression model is measured by δz_i. If $\delta > 0$, the probability of $p_i = 1$, that is, of the ith team winning, increases. On the other hand, if $\delta < 0$, there is a greater probability of a loss. The symmetric logit model can be considered as a particular case of the previous model when $\delta = 0$. In this case, the likelihood function is obtained as

$$l(y|x, \beta, \delta) = \prod_{i=1}^{n} \int_0^{\infty} [F(x_i'\beta + \delta z_i)]^{y_i}[1 - F(x_i'\beta + \delta z_i)]^{1-y_i} g(z_i)dz_i. \tag{3}$$

We assume that the prior distribution of the coefficients is normal and noninformative, that is, $\beta_j \sim N(0, \sigma_j^2), \forall j = 1, \dots, k$, and $\delta \sim N(0, \sigma_\delta^2)$, considering $\sigma_j > 0, \forall j = 1, \dots, k$, and $\sigma_\delta > 0$ sufficiently large, and noting the absence of prior knowledge about the parameters of interest, which facilitates comparison with the frequentist model. The values of the variances considered are $\sigma_j^2 = 10^8, \forall j = 1, \dots, k$, and $\sigma_\delta^2 = 10^8$.

By combining these prior assumptions with the likelihood shown in (3), we obtain the posterior distribution for the parameters β and δ:

$$\begin{aligned} \pi(\beta, \delta | y, x) &\propto l(y|x, \beta, \delta)\pi(\beta, \delta) \\ &= \left\{ \prod_{i=1}^{n} \int_0^{\infty} [F(x_i'\beta + \delta z_i)]^{y_i}[1 - F(x_i'\beta + \delta z_i)]^{1-y_i} g(z_i)dz_i \right\} \pi(\beta, \delta), \end{aligned}$$

where $\pi(\beta, \delta)$ is the prior distribution of (β, δ).

(β, δ) can be sampled from this posterior distribution using the WinBUGS package,[1] based on Gibbs sampling applying Markov Chain Monte Carlo (MCMC) methods (see Carlin & Polson, 1992, and Gilks, Richardson, & Spiegelhalter, 1995, for further details).

2.3 | Orthogonal regression

Econometric methods provide a basis for modifying, refining, or refuting theoretical conclusions, by obtaining signs and magnitudes of the variables to be related. For this purpose, the coefficients of the model are obtaiden in the following way. The model for k independent variables and n

[1]Windows Bayesian inference using Gibbs Sampling, developed jointly by the MRC Biostatistics Unit (University of Cambridge, Cambridge, UK) and the Imperial College School of Medicine at St. Mary's, London (Lunn, Thomas, Best, & Spiegelhalter, 2000).

$$y = x_j'\beta + u, \tag{4}$$

observations, and assuming that the random disturbance, $u = (u_1, u_2, \dots, u_n)'$, is centred, homoscedastic, and uncorrelated. In the usual method of estimation, by ordinary least squares (OLS), x_j' must be linearly independent (i.e., none of the regressors can be written as exact linear combinations of the remaining regressors in the model) because otherwise nonunique (perfect multicollinearity) or highly unstable (near multicollinearity) solutions may be obtained. Other problems caused by the presence of near multicollinearity among the independent variables include inflated variances and covariances, inflated correlations, and inflated prediction variance. In practice, it is almost impossible to find two or more variables that may not be correlated. Therefore, when this problem occurs, as in the present analysis, various solutions must be considered, such as using prior information, combining cross-sectional and time series data, transforming the variables, increasing the size of the sample (with additional data) or eliminating one or more variables. Usually, the last option is the only one applicable, and so alternative estimation methods must be applied. Macedo (2017) proposed a list of methods with which to estimate (4) when the degree of multicollinearity is severe, including ridge regression, principal component regression, partial least squares regression, continuum regression, lasso, elastic net, least angle regression, and generalised maximum entropy. However, in this paper, we apply orthogonal regression, a methodology that was introduced by Novales, Salmerón, García, García & López (2015) for the regression model (4) with three independent variables and which was later expanded by Salmerón, García, García, and García (2016) for four variables. This approach is based on replacing one of the exogenous variables in the regression model by the components that are not explained by the remaining exogenous variables. That is, instead of analysing the model

$$y = f(x_1, x_2, \dots, x_j, \dots, x_k) + u, \tag{5}$$

we work with the model

$$y = f(x_1, x_2, \dots, e_j, \dots, x_k) + v, \tag{6}$$

where v is a vector of random disturbances and e_j are the residuals obtained after estimating the following model by OLS:

$$x_j = x_{-j}'\alpha + w,$$

where x_{-j}' coincides with x_j' after eliminating x_j and w is a vector of random disturbances. Note that e_j is orthogonal to the variables in x_{-j}', and therefore, the degree of collinearity in (6) is less than in (5). Accordingly, the multicollinearity problem may be alleviated or even eliminated (if the process is repeated $k-2$ times) by repeating this process as many times as necessary. The following interesting questions arise concerning the orthogonal regression:

- The residuals of models (5) and (6) are the same, and so the joint significance and estimation of the variance of the random disturbance will coincide.
- The estimation and inference with the orthogonal variable is the same in both models. In other words, the estimation of the coefficient of the orthogonalised variable and its experimental value in the test of individual significance, in the initial model and in the model with orthogonal variables, will coincide.[2] Nevertheless, its interpretation in model (6) is modified: Changes are produced in y due to the variation of x_j, unrelated to other exogenous variables.

[2]For example, if $y = a + bx + u$ and $y = c + dxx + v$, where xx is the orthogonalisation of x, then the estimation of b and d by OLS is the same as the experimental value of the individual significance.

- Estimations of the coefficients of other exogenous variables in x'_{-j} coincide with those obtained from the model $y = x'_{-j}\gamma + \varepsilon$.

Finally, the orthogonal variable can be chosen according to the following criteria:

- Because the interpretation of the orthogonal variable changes, choose a variable considered less important.
- Because the interpretation of the orthogonal variable changes, choose a variable that facilitates interpretation with the original model and which is of interest to the researcher.
- Choose the variable with the highest variance inflation factor (VIF).[3]
- Because the inference of the orthogonal variable does not change, choose a variable whose coefficient is statistically significant.

3 | DATABASE AND SELECTION OF THE VARIABLES

This section describes the process used to select the variables that are part of the econometric model analysed in Section 4. The original database (the variables shown in Tables A1 and A2)[4] was downloaded from http://www.acb.com/. The glossary of all the variables considered in this paper can be consulted at http://www.basketball-reference.com/about/glossary.html. In accordance with the indications shown in this website and the recommendations of Kubatko, Oliver, Pelton and Rosenbaum (2007), we computed a set of statistical variables related to the performance of the ACB teams Real Madrid and Barcelona F.C. (see Tables A3 and A4). Finally, from the variables that presented significantly different coefficients for wins and losses, we selected those that best reflected the characteristics of each team, and which therefore enabled us to develop a model with good fit. In this process, a very important role was played by the careful monitoring of each team during the season.

3.1 | Real Madrid

In this section, we analyse the information generated in the 80 games played by the Real Madrid basketball team (RMCF) in the 2013–2014 season in official competitions. Tables A7 and A8 show the ratings for all these games, according to the final result (win or lose), match site (local or visitor) and competition. Significantly, there were only 12 losses and only two of these were at home. The chi-square test of independence shows that winning or losing and playing at home or away are dependent variables ($p = .026$). Moreover, the same test shows that win–lose and type of competition are independent variables ($p = .704$).

Tables A1 and A3 show the average values obtained (total, win and lose) for a set of variables computed for RMCF and its opponents. The values that are statistically different at 5% significance (equality of the means test) are highlighted. The following variables were selected, according to predictability, for the econometric model:

Q1: points scored by RMCF in the first quarter.
T1I: free throws for the opponent.

[3]The VIF is one of the most commonly used methods to detect multicollinearity. For more details, see, for example, García, García, López, and Salmerón (2015).
[4]The tables referenced in this section are in Appendix A. In addition, the meanings of the acronyms are presented in Tables A5 and A6.

PR: points received by RMCF.

RT: total rebounds (%) captured by RMCF.

Another factor taken into account was a variable describing whether a particular RMCF player, Mirotic, played well on a given occasion. This player was chosen because he was the second highest points scorer, captured the highest number of rebounds, and was the highest rated player (on average, per game). Thus, he was of crucial importance to RMCF during the season in question.[5] For the purposes of our study, this player was considered to have achieved a high level of performance when in a game, he exceeded his seasons' average score (average of 0.25 points/possession).

3.2 | F.C. Barcelona

The Barcelona basketball team (FCB) played 79 official games in the 2013–2014 season. Tables A9 and A10 show the ratings for all games according to the final result (win or lose), match site (local or visitor), and competition. In this case, the chi-square test of independence shows that win–lose, playing at home or away, and the type of competition are all independent variables ($p = .109$ and $p = .774$, respectively). In relation to the team performance, Tables A2 and A4 show the average values (total, win and lose) for several variables for FCB and its opponents. The values that are statistically different at 5% significance (according to the equality of means test) are highlighted. The very large set of variables obtained related to the percentage of throws that are significant, highlighting the importance of the three-point line in basketball today, are summarised by means of a dummy variable, T3, which takes the value 1 if FCB achieved 30% or more of its score in three-point throws, and 0 otherwise). The quarter-by-quarter statistical analysis revealed a significant situation in the fourth quarter. As can be seen in Table A11, if the team is winning at the start of this quarter, there is a statistically significant association with the result ($p < .0001$). In accordance with the information derived from these tables, the following variables in the econometric model predict the probability of a win:

Q4W: dummy variable that takes the value 1 if FCB is winning at the start of the fourth quarter, and 0 otherwise.

PR: points received by FCB.

T3: dummy variable that takes the value 1 if FCB achieves 30% or more of its score from three-point throws, and 0 otherwise.

PACE: number of offensive plays by the opponent.

RO: percentage of offensive rebounds captured.

Finally, the performance of the FCB player Tomic was analysed, taking into account that this player is highly important to the team. This player in particular was chosen for analysis because he heads several of the team's statistical categories, despite being only fourth in the ranking of minutes played. He is the team's most highly rated player and the highest scorer. In addition, he takes the most field throws and free throws, makes most blocks, captures most rebounds (defensive and offensive), and receives most fouls. Furthermore, his scoring performance (a dichotomous variable distinguishing whether the player exceeds his average scoring rate per minute of play) shows

[5]Mirotic's statistics for 2013/14: 24.2 minutes played per game, 13.09 points scored per game, 5.3 rebounds captured per game, and player rating of 15.62 per game.

indications of association with the team's score ($p = .26$).[6] In this case, his average score for the season was 0.26 points/possession.

4 | ESTIMATING THE MODELS

In this section, we compare the noninformative Bayesian estimation of a logit model and the frequentist estimation of the logit model, for RMCF and FCB, seeking to show that the two models provide similar results in terms of parameter estimates and fit. These results are then compared with the Bayesian skewed estimate of the logit model, and we show that this new approximation considerably improves the overall fit and prediction. To assess the quality of fit and prediction for the standard logit model and for the Bayesian models analysed, we propose three different measures: (a) the percentage of correct fits obtained considering the estimation sample, that is, the 2013–2014 season; (b) the percentage of correct predictions obtained considering the 2014–2015 season; (c) a statistical fit measure such as the Akaike information criterion (AIC) or the deviance information criterion (DIC). The posterior distributions for Bayesian models were simulated using WinBUGS. In total, 500,000 iterations were carried out (after a burn-in period of 100,000 simulations) for each of the teams. Three different chains were performed, and the convergence was evaluated for all parameters using tests provided within the WinBUGS Convergence Diagnostics and Output Analysis software.

4.1 | Frequentist and standard Bayesian estimation of the logistic regression model

4.1.1 | RMCF

The results obtained by estimating the logistic model from the frequentist and noninformative Bayesian standpoints are given in Table 1. As expected, because the prior information is noninformative, the Bayesian estimations were similar to those obtained by the standard frequentist model. The variables e_{PR} and e_{RT} were obtained by considering the multicollinearity correction procedure developed in Section 2.3, and its interpretation provides more detailed results.[7] As can be seen in this table, the coefficient of e_{PR} is significant at 1%, whereas those of *Mirotic* and *T1I* are significant at 5% and those of e_{RT} and *Q1*, at 10%. T1I and e_{PR} are both inversely related to a high probability of winning. If the opposing team attempts one additional free throw, the probability of winning (relative to losing) decreases by 58.8 % (odds ratio of 0.412). The more two- and three-point goals received during the second, third, and fourth quarters and the absence of Mirotic (the interpretation of e_{PR}) decrease the relative probability of winning by 35.9 % (odds ratio of 0.641). If Mirotic plays a good game, the probability of winning (with respect to losing) is greatly increased. A higher number of points scored by RMCF in the first quarter increases the relative probability of winning. Finally, the total number of rebounds obtained without Mirotic (or when Mirotic plays a “bad” game) during the second, third, and fourth quarters (the interpretation of e_{RT}) increases the relative probability of winning.

For the RMCF database, we obtained an AIC of 24.56 for the frequentist logit model and a DIC of 24.23 for the Bayesian logit model. The two models produced the same percentages of correct

[6]Tomic's statistics for the 2013–2014 season were, on average, 20.8 min played, 10.9 points scored, and 8 rebounds captured, for a player rating of 15.42 per game.

[7]PR and RT variables were chosen because they presented the highest VIFs: VIF(PR) = 41.29 and VIF(RT) = 39.50.

TABLE 1 Frequentist and noninformative Bayesian estimations (RMCF)

	Frequentist			Noninformative Bayesian		
Variables	$\hat{\beta}$	***SD***	***p* value**	$\hat{\beta}$	**MC error**	***SD***
Intercept	−3.481	4.500	0.439	−3.846	0.087	3.832
Q1	1.173	0.638	0.066*	1.552	0.022	0.549***
T1I	−0.886	0.389	0.023**	−1.196	0.017	0.432***
Mirotic	8.130	3.563	0.023**	10.65	0.136	3.85***
e_{PR}	−0.445	0.154	0.004***	−0.619	0.008	0.228***
e_{RT}	0.429	0.228	0.060*	0.549	0.008	0.243***
	AIC = 24.56			DIC = 24.23		
% correct fit	96.25			96.25.		
% correct prediction	65.38			65.38.		

Note. SD = standard deviation; MC Error = Markov Chain Error. * This indicates 10% significance level. ** This indicates 5% significance level. *** This indicates 1% significance level.

fits and predictions. Table 1 shows that the accuracy, that is, the proportions of wins that were correctly classified by the models, is around 96.25% for the fit (2013–2014 season) and 65.38% for the prediction (2014–2015 season). The threshold probabilities used to fit and predict a win corresponded to the sample frequencies of wins, that is, 0.85 and 0.82, respectively.

4.1.2 | FCB

Table 2 shows the estimation results for FCB obtained using frequentist and noninformative Bayesian estimation methods. Again, the results are similar for both procedures. When Tomic plays, when the team is winning at the start of the fourth quarter and when the team scores over 30% of its total points as three-point goals are all significant, positive factors in the probability of winning. On the other hand, the more points received significantly decreases the probability of winning. In both cases, the coefficients of the two variables representing the orthogonal residues (e_{PACE} and e_{RO}) are not statistically significant.[8] An AIC of 43.14 was obtained for the frequentist logit model, and a DIC of 42.59 for the Bayesian logit model. Again, the two models produced the same percentages of correct fits and predictions. The proportion of wins that were correctly classified by the models was around 92.11% for fits and 82.89% for predictions. The threshold probabilities again corresponded to the sample frequencies of winning, that is, 0.77 and 0.70, respectively.

4.2 | Asymmetric Bayesian estimation of the logistic regression model

4.2.1 | RMCF

Table 3 summarises the results of the Bayesian estimation for RMCF, including a variable to measure the possible asymmetry of the model. This model obtains the same results as the frequentist and uninformative Bayesian ones in terms of significant factors. However, the estimated coefficients differ considerably from those of the previous two RMCF models, although the signs remain the same. For example, in this case, if the opposing team attempts more free throws, the probability of winning (relative to losing) decreases by 95.5% (odds ratio of 0.05). Moreover, the coefficient

[8]In this case, PACE and RO variables were those that presented the highest VIFs: VIF(PACE) = 99.26 and VIF(RO) = 10.86.

TABLE 2 Frequentist and noninformative Bayesian estimations (FCB)

	Frequentist			Noninformative Bayesian		
Variables	$\hat{\beta}$	***SD***	***p* value**	$\hat{\beta}$	**MC error**	***SD***
Intercept	9.179	4.867	0.059*	9.158	0.171	4.794***
Q4W	3.618	1.317	0.006***	4.836	0.015	1.48***
PR	−0.172	0.065	0.008***	−0.187	0.002	0.067***
T3	3.260	1.489	0.029**	3.99	0.019	1.409***
Tomic	2.892	1.175	0.014**	3.571	0.029	1.456***
e_{PACE}	−0.076	0.148	0.609	−0.105	0.0008	0.159
e_{RO}	0.058	0.476	0.223	0.074	0.0004	0.082
	AIC = 43.14			DIC = 42.59		
% correct fit	92.11			92.11.		
% correct prediction	82.89			82.89.		

Note. SD = standard deviation; MC Error = Markov Chain Error. * This indicates 10% significance level. ** This indicates 5% significance level. *** This indicates 1% significance level.

TABLE 3 Asymmetric Bayesian logit estimation results (RMCF)

Variables	$\hat{\beta}$	**MC error**	***SD***
Intercept	−0.204	0.22	8.187
Q1	3.717	0.068	1.398***
T1I	−2.897	0.049	1.046***
Mirotic	18.79	0.168	6.313***
e_{PR}	−1.541	0.025	0.596***
e_{RT}	1.388	0.024	0.759***
δ	−6.076	0.691	13.03

Note. SD = standard deviation; MC Error = Markov Chain Error. Dependent variable: Win. $N = 79$. DIC = −7.23. % correct fit = 96.25. % correct prediction = 70.51. * This indicates 10% significance level. ** This indicates 5% significance level. *** This indicates 1% significance level.

of variable δ, which measures the asymmetry of the data, is statistically irrelevant, indicating that the marked skewness of the data is well captured by the previous frequentist and uninformative Bayesian estimation models. The DIC measure is equal to −7.23 for the skewed link model, that is, it is notably lower than the values obtained by the nonskewed models. This major reduction in the DIC measure indicates a significant increase in the level of fit. Furthermore, improved prediction results are obtained by the skewed model. The percentage of correct predictions is 70.51%, which means that the prediction leverage of this model is much better than with the symmetric models. The percentage of correct fitting is again 96.25%. Thus, even though the variable measuring the asymmetry is not significant, the Bayesian estimation of the asymmetric model achieves better results, both for fit and for prediction.

4.2.2 | FCB

Table 4 shows the results of the asymmetric Bayesian estimation for FCB. In this case, in comparison with the two previous symmetric models, we obtain the same positive and negative relevant factors, that is, the presence of Tomic, starting the fourth quarter ahead on the scoreboard, scoring a high proportion of three-point goals and receiving less points are all relevant factors to the

TABLE 4 Asymmetric Bayesian logit estimation results (FCB)

Variables	$\hat{\beta}$	MC error	*SD*
Intercept	28.61	1.209	26.71
Q4W	46.64	0.513	15.77***
PR	−1.399	0.021	0.447***
T3	26.13	0.319	12.19***
Tomic	30.39	0.353	13.01***
e_{PACE}	0.645	0.016	1.303
e_{RO}	0.479	0.009	0.847
δ	47.27	0.469	15.8***

Note. SD = standard deviation. Dependent variable: Win. $N = 76$. DIC = 6.12. % correct fit = 100. % correct prediction = 82.89. * This indicates 10% significance level. ** This indicates 5% significance level. *** This indicates 1% significance level.

probability of a win by FCB. However, the δ coefficient is significant and positive at 1% of significance. This variable, therefore, adjusts the estimated probability of a win, that is, it increases the probability of a win and, perhaps, counteracts the evident asymmetry in the data. Again, the estimated coefficients differ considerably from those of the previous two FCB models. This difference is further accentuated in the estimation of the intercept. We believe that in the first two models, the estimated intercept may contain part of the asymmetry effect made apparent in the asymmetric model. The DIC measure is equal to 6.12, which improves upon the results obtained from the uninformative symmetric Bayesian estimation (DIC = 42.59). This model provides a fit of 100% and a prediction value of 82.89%. Thus, it greatly improves on the fit and equals the prediction results with respect to the results obtained by the frequentist and uninformative symmetric Bayesian models. Obviously, these outcomes are explained by the increase in the probability of fitting the cases in which $y_i = 1$, induced by the skewed model, because δ was positive.

5 | CONCLUSIONS

In this paper, a novel statistical method, that of asymmetric logistic regression, is presented to extend the range of quantitative tools available to the analyst. The advantage of this regression model over the standard version is that the skewed link in the dependent variable is taken into account. This case is presented in the study, in which two basketball teams have a very unbalanced number of wins and losses. We apply asymmetric logistic regression to analyse the influence of certain factors on the probability of the team winning its games. To our knowledge, this tool has not been previously applied in the analysis of sports performance. With this new method, the model detects higher levels of significance in the factors relevant to the team's wins and losses. Furthermore, the method improves the percentage of correct fits and predictions obtained even when the asymmetry coefficient is not significant (RMCF). Thus, application of this approach could enable coaches to simulate games more efficiently, taking relevant factors into account and estimating/predicting the probability of sporting success.

In view of the results obtained, various practical observations can be made to assist coaches in their decision, for RMCF, regarding the importance of the first quarter, the number of rebounds captured, and the number of personal fouls committed, and for FCB, the importance of three-point goals and of being ahead on the scoreboard at the start of the fourth quarter. Clearly, if coaches wish to improve their teams' performance, they should manage the resources available in such a way as to maximise the probability of a win, by paying special attention to the key factors identified. The findings of this study may be profitably employed not only by the coaches of RMCF and FCB but also by those of their opponents, who may thus design better training programmes and strategies. In conclusion, the approach shown in this study and the results obtained provide coaches with valuable data for the efficient management of their teams.

ACKNOWLEDGEMENTS

This work was partially funded by grants ECO2013–47092 (Ministerio de Economía y Competitividad, Spain), ECO2017–85577–P (Ministerio de Economía, Industria y Competitividad, Agencia Estatal de Investigación, Spain) and Quantitative Methods in Economic and Business Department from University of Granada.

ORCID

José María Pérez-Sánchez http://orcid.org/0000-0002-7491-4345

REFERENCES

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679. doi:https//doi.org/10.1080/01621459.1993.10476321.

Baghal, T. (2012). Are the "Four Factors" indicators of one factor? An application of structural equation modeling methodology to NBA data in prediction of winning percentage. *Journal of Quantitative Analysis in Sports*, *8*(1), 1–17. doi:https//doi.org/10.1515/1559-0410.1355.

Bar-Eli, M., & Tractinsky, N. (2000). Criticality of game situations and decision making in basketball: An application of performance crisis perspective. *Psychology of Sport and Exercise*, *1*, 27–39. doi:https//doi.org/10.1016/\ignorespacesS1469-0292(00)00005-4.

Bermúdez, L. L., Pérez, J. M., Ayuso, M., Gómez, E., & Vázquez, F. J. (2008). A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics*, *42*, 779–786. doi:https//doi.org/10.1016/j.insmatheco.2007.08.002.

Casals, M., & Martínez, J. A. (2013). Modelling player performance in basketball through mixed models. *International Journal of Performance Analysis in Sport*, *13*(1), 64–82.

Carlin, B. P., & Polson, N. G. (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. *Bayesian Statistics*, *4*, 577–86.

Chen, M. H., Dey, D. K., & Shao, Q. M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, *94*, 1172–1186. doi:https//doi.org/10.1080/01621459.1999.10473872.

Csapo, P., Avugos, S., Raab, M., & Bar-Eli, M. (2015). How should "hot" players in basketball be defended? The use of fast–and–frugal heuristics by basketball coaches and players in response to streakiness. *Journal of Sports Sciences*, *33*(15), 1580–1588. doi:https//doi.org/10.1080/02640414.2014.999251.

De La Torre–Ruiz, J. M., & Aragón–Correa, J. A. (2012). Performance of newcomers in highly interdependent teams: The case of basketball teams. *European Sport Management Quarterly*, *12*(3), 205–226. doi:https//doi.org/10.\ignorespaces1080/16184742.2012.679287.

Del Corral, J., García–Unanue, J., & Herencia, F. (2016). Are NBA policies that promote long-term competitive balance effective? What is the price?. *The Open Sports Sciences Journal*, *19*, 15–27.

Del Corral, J., Maroto, A., & Gallardo, A. (2017). Are former professional athletes and native better coaches? Evidence from Spanish basketball. *Journal of Sports Economics*, *18*(7), 698–719. doi:https//doi.org/10.1177/1527002515595266.

Fizel, J. L., & D'Itri, M. P. (1996). Estimating managerial efficiency: The case of college basketball coaches. *Journal of Sport Management*, *10*, 435–445.

García, C. B., García, J., López, M. M., & Salmerón, R. (2015). Collinearity: Revisiting the variance inflation factor in ridge regression. *Journal of Applied Statistics*, *42*(3), 648–661. doi:https//doi.org/10.1080/02664763.2014.980789.

García, J., Ibáñez, S., Martínez De Santos, R., Leite, N., & Sampaio, J. (2013). Identifying basketball performance indicators in regular season and playoff games. *Journal of Human Kinetics*, *36*(1), 161–168. doi:https//doi.org/10.2478/hukin-2013-0016.

García, J., Ibáñez, S., Góez, M. A., & Sampaio, J. (2014). Basketball game-related statistics discriminating ACB league teams according to game location, game outcome and final score differences. *International Journal of Performance Analysis in Sport*, *14*(2), 443–452.

Giambatista, R. C. (2004). Jumping through hoops: A longitudinal study of leader life cycles in the NBA. *The Leadership Quarterly*, *15*, 607–624. doi:https//doi.org/10.1016/j.leaqua.2004.07.002.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1995). Introducing Markov Chain Monte Carlo. In Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.), *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Gómez, M., Lorenzo, A., Sampaio, J., & Ibáñez, S. J. (2006). Differences in game-related statistics between winning and losing teams in women?s basketball. *Journal of Human Movement Studies*, *51*, 357–369. doi:https//doi.org/10.2466/PMS.106.1.43-50.

Gómez, M. A., Lorenzo, A., Ortega, E., Sampaio, J., & Ibáñez, S. J. (2009). Game related statistics discriminating between starters and nonstarters players in Women's National Basketball Association League (WNBA). *Journal of Sports Science and Medicine*, *8*(2), 278–283.

Gómez, M. A., Lorenzo, A., Ibáñez, S. J., & Sampaio, J. (2013). Ball possession effectiveness in men's and women's elite basketball according to situational variables in different game periods. *Journal of Sports Sciences*, *31*(14), 1578–1587. doi:https//doi.org/10.1080/02640414.2013.792942.

Gómez, M. A., Battaglia, O., Lorenzo, A., Lorenzo, J., Jiménez, S., & Sampaio, J. (2015). Effectiveness during ball screens in elite basketball games. *Journal of Sports Sciences*, *33*(17), 1844–1852. doi:https//doi.org/10.1080/02640414.2015.1014829.

Gómez, S., & Salmerón, R. (2015). Life cycles or longer tenures? a performance and employment duration model for Spanish basketball coaches. *Coaching: An International Journal of Theory, Research and Practice*, *8*(1), 36–52. doi:https//doi.org/10.1080/17521882.2014.993672.

Gómez, S., & Salmerón, R. (2016). Recovering performance in the short term after coach succession in Spanish basketball organisations. *Coaching: An International Journal of Theory, Research and Practice*, *9*(1), 24–37. doi:https//doi.org/10.1080/17521882.2015.1119169.

Ibáñez, S. J., Sampaio, J., Feu, S., Lorenzo, A., Gómez, M. A., & Ortega, E. (2008). Basketball game–related statistics that discriminate between teams' season-long success. *European Journal of Sport Science*, *8*(6), 369–372. doi:https//doi.org/10.1080/17461390802261470.

Karipidis, A., Fotinakis, P., Taxildaris, K., & Fatouros, J. (2001). Factors characterizing a successful performance in basketball. *Journal of Human Movement Studies*, *41*, 385–397.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, *3*(3), 1–24. doi:https//doi.org/10.2202/1559-0410.1070.

Lewis, M. (2004). *Moneyball: The Art of Winning an Unfair Game*. New York: Norton.

Lorenzo, A., Gómez, M. A., Ortega, E., Ibáñez, S. J., & Sampaio, J. (2010). Game related statistics which discriminate between winning and losing under-16 male basketball games. *Journal of Sports Science and Medicine*, *9*(4), 664–668.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–37. doi:https//doi.org/10.1023/A:\ignorespaces1008929526011.

Macedo, P. (2017). Ridge regression and generalized maximum entropy: An improved version of the Ridge-GME Parameter Estimator. *Communications in Statistics–Simulation and Computation*, *46*(5). doi:https//doi.org/10.1080/\ignorespaces03610918.2015.1096378.

Martínez, J. A., & Martínez, L. (2010). A probabilistic method to statistically classify players in basketball. *Revista Internacional de Ciencias del Deporte*, *6*(18), 13–36.

Martínez, J. A., & Caudill, S. B. (2013). Does midseason change of coach improve team performance? Evidence from the NBA. *Journal of Sport Management*, *27*, 108–113.

McGoldrick, K., & Voeks, L. (2004). We got game!: An analysis of win/loss probability and efficiency differences between the NBA and WNBA. *Journal of Sports Economics*, *6*(1), 5–23. doi:https//doi.org/10.1177/1527002503262649.

Mikolajec, K., Maszczyk, A., & Zajac, T. (2013). Game indicators determining sports performance in the NBA. *Journal of Human Kinetics*, *37*, 145–151. doi:https//doi.org/10.2478/hukin-2013-0035.

Novales, A., Salmerón, R., García, C., García, J., & López, M. (2015). Tratamiento de la multicolinealidad aproximada mediante variables ortogonales. *Anales de Economía Aplicada*, 1212–1227.

Oliver, D. (2004). *Basketball on Paper Rules and Tools for Performance Analysis*. Washington DC: Brassey's.

Padulo, J., Attene, G., Migliaccio, G. M., Cuzzolin, F., Vando, S., & Paolo Ardigò, L. (2015a). Metabolic optimisation of the basketball free throw. *Journal of Sports Sciences*, *33*(14), 1454–1458. doi:https//doi.org/10.1080/02640414.2014.990494.

Padulo, J., Laffaye, G., Haddad, M., Chaouachi, A., Attene, G., Migliaccio, G. M., Chamari, K., & Pizzolato, F. (2015b). Repeated sprint ability in young basketball players: One vs. two changes of direction (Part 1). *Journal of Sports Sciences*, *33*(14), 1480–1492. doi:https//doi.org/10.1080/02640414.2014.992936.

Pérez–Sánchez, J. M., Negrín–Hernández, M. A., García–García, C., & Gómez–Déniz, E. (2014). Bayesian asymmetric logit model for detecting risk factors in motor ratemaking. *Astin Bulletin*, *44*(2), 445–457. doi:https//doi.org/10.1017/asb.2013.32.

Sáez–Castillo, A. J., Olmo–Jiménez, M. J., Pérez, J. M., Negrín, M., Arcos, A., & Díaz, J. (2010). Bayesian analysis of nosocomial infection risk and length of stay in a department of general and digestive surgery. *Value in Health*, *13*(4), 431–439. doi:https//doi.org/10.1111/j.1524-4733.2009.00680.$$\backslash$times$.

Salmerón, R., García, C. B., García, J., & García, C. (2016). Tratamiento de la colinealidad mediante la regresión con variables ortogonales: Un ejemplo económico. *Boletín de Estadística e Investigación Operativa*, *32*(3), 184–202.

Sampaio, J., Drinkwater, E. J., & Leitea, N. M. (2010). Effects of season period, team quality, and playing time on basketball players' game-related statistics. *European Journal of Sport Science*, *10*(2), 141–149. doi:https//doi.org/10.1080/17461390903311935.

Sampaio, J., & Janeira, M. (2003). Statistical analyses of basketball team performance: Understanding teams' wins and losses according to a different index of ball possessions. *International Journal of Performance Analysis in Sport*, *3*(1), 40–49.

Sampaio, J., Janeira, M., Ibáñez, S., & Lorenzo, A. (2006). Discriminant analysis of game-related statistics between basketball guards, forwards and centres in three professional leagues. *European Journal of Sport Science*, *6*, 173–178.

Sampaio, J., Lago–Peñas, C., & Drinkwater, E. (2008). Explanations for the United States of America's dominance in basketball at the Beijing Olympic Games. *Journal of Sports Sciences*, *28*(2), 147–52.

Sánchez, J. M., Castellanos, P., & Dopicoa, J. A. (2007). The winning production function: Empirical evidence from Spanish basketball. *European Sport Management Quarterly*, *7*(3), 283–300. doi:https//doi.org/10.1080/\ignorespaces16184740701511177.

StataCorp (2015). *Stata Statistical Software: Release 14 College Station*. TX: StataCorp LP.

Stekler, H. O., & Klein, A. (2012). Predicting the outcomes of NCAA basketball championship games. *Journal of Quantitative Analysis in Sports*, *8*(1), 1–10. doi:https//doi.org/10.1515/1559-0410.1373.

Stoszkowski, J., & Collins, D. (2016). Sources, topics and use of knowledge by coaches. *Journal of Sports Sciences*, *34*(9), 794–802. doi:https//doi.org/10.1080/02640414.2015.1072279.

Stukel, T. (1988). Generalized logistic model. *Journal of the American Statistical Association*, *83*(402), 426–431. doi: https//doi.org/10.1080/01621459.1988.10478613.

Summers, M. R. (2013). How to win in the NBA playoffs: A statistical analysis. *American Journal of Management*, *13*(3), 11–24.

Teramoto, M., & Cross Chad, L. (2010). Relative importance of performance factors in winning NBA games in regular season versus playoffs. *Journal of Quantitative Analysis in Sports*, *6*(3), 1–19.

Viggiano, A., Chieffi, S., Tafuri, D., Messina, G., Monda, M., & De Luca, B. (2014). Laterality of a second player position affects lateral deviation of basketball shooting. *Journal of Sports Sciences*, *32*(1), 46–52. doi:https//doi.org/10.1080/02640414.2013.805236.

How to cite this article: Pérez-Sánchez JM, Salmerón-Gómez R, Ocaña-Peinado FM. A Bayesian asymmetric logistic model of factors underlying team success in top-level basketball in Spain. *Statistica Neerlandica*. 2018;1–22. https://doi.org/10.1111/stan.12127

APPENDIX A

TABLE A1 Boxscore stats means for RMCF and its opponent depending on game result

Variable	RMCF Lose	RMCF Win	RMCF Total	RMCF opponent Lose	RMCF opponent Win	RMCF opponent Total
Points	**80.500**	**88.265**	87.100	**88.583**	**71.309**	73.900
T2C	**18.833**	**21.691**	21.263	21.417	19.941	20.163
T2I	39.583	38.294	38.488	38.333	40.500	40.175
T2 %	**47.433**	**56.864**	55.450	**56.798**	**49.430**	50.535
T3C	8.083	9.559	9.338	**9.333**	**6.912**	7.275
T3I	23.667	24.662	24.513	23.167	22.015	22.188
T3 %	34.204	38.570	37.915	**40.436**	**31.373**	32.733
TCC	**26.917**	**31.250**	30.600	**30.750**	**26.853**	27.438
TCI	63.250	62.956	63.000	61.500	62.515	62.363
TC %	**42.505**	**49.604**	48.540	**50.343**	**43.118**	44.202
T1C	18.583	16.206	16.563	**17.750**	**10.691**	11.750
T1I	**24.167**	**20.147**	20.750	**22.500**	**14.971**	16.100
T1 %	77.273	81.000	80.441	77.590	72.443	73.215
RT	32.833	35.574	35.163	**37.083**	**32.603**	33.275
RD	**22.250**	**25.632**	25.125	**27.167**	**21.926**	22.713
RO	10.583	9.941	10.038	9.917	10.676	10.563
Asis	**14.000**	**18.603**	17.913	17.917	14.485	15.000
BR	**6.417**	**9.206**	8.788	4.917	5.500	5.413
BP	10.750	10.956	10.925	**10.083**	**14.515**	13.850
FPC	**22.500**	**18.294**	18.925	**24.250**	**21.956**	22.300
FPR	24.000	21.853	22.175	**22.083**	**18.206**	18.788
PQ	48.833	50.397	50.163	**56.167**	**43.853**	45.700
PI	31.333	35.574	34.938	**37.167**	**30.044**	31.113
Val	**83.500**	**110.132**	106.138	**98.167**	**64.147**	69.250
Q1	**17.583**	**21.559**	20.963	**22.000**	**17.529**	18.200
Q2	20.583	22.588	22.288	21.167	18.250	18.688
Q3	20.000	21.765	21.500	**22.417**	**16.662**	17.525
Q4	21.250	21.897	21.800	20.667	18.603	18.913
Q1 - Q1 (R)	**−4.417**	**4.029**	2.763			
Q2 - Q2 (R)	−0.583	4.338	3.600			
Q3 - Q3 (R)	**−2.417**	**5.103**	3.975			
Q4 - Q4 (R)	0.583	3.294	2.888			

TABLE A2 Boxscore stats means for FCB and its opponent depending on game result

Variable	FCB Lose	FCB Win	FCB Total	FCB opponent Lose	FCB opponent Win	FCB opponent Total
Points	**72.176**	**83.533**	81.026	**83.353**	**68.750**	71.974
T2C	**19.294**	**23.000**	22.182	21.882	18.533	19.273
T2I	38.176	40.217	39.766	42.824	41.233	41.584
T2 %	**50.652**	**57.222**	55.771	**51.787**	**44.842**	46.375
T3C	**6.470**	**8.702**	8.207	8.294	6.566	6.948
T3I	22.471	22.267	22.312	20.941	20.133	20.312
T3 %	28.405	38.442	36.226	**38.973**	**33.278**	34.535
TCC	**25.765**	**31.700**	30.39	30.176	25.100	26.221
TCI	60.647	62.483	62.078	63.765	61.367	61.896
TC %	**42.530**	**50.830**	48.997	**47.390**	**40.941**	42.364
T1C	**14.176**	**11.433**	12.039	14.706	11.983	12.584
T1I	20.765	16.117	17.143	18.765	15.85	16.494
T1 %	69.728	70.142	70.051	77.695	75.749	76.179
RT	35.176	36.750	36.403	**35.235**	**31.401**	32.247
RD	24.882	26.317	26.000	**26.353**	**21.333**	22.442
RO	10.294	10.433	10.403	8.882	10.067	9.805
Asis	**15.118**	**18.233**	17.545	**16.176**	12.650	13.429
BR	**4.588**	**6.416**	6.013	6.882	6.433	6.532
BP	12.412	11.600	11.779	**9.588**	**12.117**	11.558
FPC	21.471	19.317	19.792	21.941	19.517	20.052
FPR	21.824	19.633	20.117	21.294	19.766	20.103
PQ	46.353	46.467	46.442	48.176	45.600	46.169
PI	**30.294**	**36.217**	34.909	**35.471**	**30.467**	31.571
Val	**73.412**	**99.183**	93.494	**93.882**	**65.950**	72.117
Q1	19.941	20.933	20.714	**20.059**	**17.233**	17.857
Q2	**16.647**	**20.600**	19.727	19.647	17.150	17.701
Q3	**17.529**	**20.533**	19.870	**21.824**	**16.767**	17.883
Q4	**18.118**	**21.400**	20.675	**21.824**	**17.750**	18.649
Q1 - Q1 (R)	**−0.118**	**3.700**	2.763			
Q2 - Q2 (R)	−3.000	**3.450**	2.026			
Q3 - Q3 (R)	−4.294	3.767	1.987			
Q4 - Q4 (R)	−3.706	3.650	2.026			

TABLE A3 Advanced stats: Mean values for RMCF and its opponent depending on game result

Variable	RMCF			RMCF opponent		
	Lose	Win	Total	Lose	Win	Total
Difference	**−8.083**	**16.956**	13.200			
% RT	**46.712**	**52.238**	51.409			
% RD	69.650	71.056	70.845	72.097	69.086	69.538
% RO	27.903	30.914	30.462	30.350	28.944	29.155
Pace	73.083	72.029	72.188	70.667	72.341	72.090
POS	83.667	81.971	82.225	80.583	83.018	82.653
Seg × POS	14.553	14.753	14.723	15.154	14.563	14.652
% BP × POS	12.765	13.334	13.249	**12.569**	**17.402**	16.677
OER	**0.962**	**1.078**	1.061	**1.101**	**0.862**	0.898
OER pot	**1.103**	**1.244**	1.223	**1.265**	**1.044**	1.077
Eficacy	**46.174**	**51.724**	50.891	**52.797**	**44.144**	45.442
% PQ	61.094	57.313	57.880	63.965	61.345	61.738
% PI	39.230	40.354	40.185	41.975	42.349	42.293
% T2I/TCI	62.581	60.640	60.931	62.132	64.876	64.465
BR/BP	0.629	0.949	0.901	0.581	0.411	0.437
Asis/BP	1.352	1.959	1.868	**2.080**	**1.174**	1.310
T1I/FPR	1.001	0.907	0.921	**1.007**	**0.799**	0.830
% T1 Ptos	**23.236**	**18.540**	19.245	**19.866**	**14.830**	15.585
% Asis × Pos	**16.694**	**22.715**	21.812	**22.382**	**17.584**	18.304
T2I × T3I	1.727	1.625	1.640	1.695	1.958	1.919
TCI × T1I	2.775	3.524	3.411	3.043	5.394	5.042
% BPnoF	0.537	0.494	0.500	0.118	0.370	0.332
eFG %	**48.885**	**57.231**	55.979	**58.008**	**48.671**	50.071
% TS	**54.428**	**61.519**	60.456	**62.436**	**51.718**	53.326

TABLE A4 Advanced stats : Mean values FCB and its opponent depending on game result

Variable	FCB Lose	Win	Total	FCB opponent Lose	Win	Total
Difference	**−11.176**	**14.783**	9.052			
% RT	**46.712**	**52.238**	51.409			
% RD	74.199	72.603	72.955	72.147	67.204	68.295
% RO	27.853	32.796	31.705	25.801	27.397	27.045
Pace	71.071	70.097	70.312	**71.976**	**69.757**	70.247
POS	81.365	80.530	80.714	80.859	79.823	80.052
Seg × POS	14.783	14.992	14.945	14.882	15.153	15.093
% BP × POS	15.267	14.380	14.576	**11.867**	**15.321**	14.558
OER	**0.888**	**1.038**	1.004	**1.031**	**0.861**	0.899
OER pot	**1.052**	**1.214**	1.178	**1.172**	**1.017**	1.052
Eficacy	**44.079**	**51.245**	49.662	**49.883**	**43.433**	44.857
% PQ	**64.322**	**56.058**	57.882	57.226	**66.655**	64.684
% PI	**38.845**	**43.317**	42.297	42.603	44.787	44.305
% T2I/TCI	63.024	64.400	64.096	67.106	67.259	67.225
BR/BP	0.429	0.617	0.576	0.760	0.586	0.624
Asis/BP	1.357	**1.795**	1.698	**1.792**	**1.161**	1.301
T1I/FPR	**0.940**	**0.810**	0.839	0.878	0.791	0.810
% T1 Points	**20.048**	**13.611**	15.032	17.846	17.317	17.434
% T3 Points	**26.419**	**30.954**	29.952	29.406	28.675	28.836
% Asis × Pos	**18.593**	**22.661**	21.763	**20.049**	**15.958**	16.861
T2I × T3I	1.872	1.955	1.937	2.172	2.234	2.220
TCI × T1I	3.313	4.446	4.196	3.706	4.735	4.508
% BPnoF	0.426	0.442	0.438	0.532	0.477	0.489
eFG %	**47.893**	**57.807**	55.618	**53.881**	**46.348**	48.011
% TS	**51.961**	**60.118**	58.317	**57.988**	**50.393**	52.070

TABLE A5 Acronyms used in Tables A1 and A2

Variable	Description
Points	Points scored
T2C	2-point field goals
T2I	2-point field goal attempts
T2 %	2-point field goal percentage
T3C	3-point field goals
T3I	3-point field goal attempts
T3 %	3-point field goal percentage
TCC	Field foals
TCI	Field goal attempts (includes both 2-point field goal attempts and 3-point field goal attempts)
TC %	Field goal percentage
T1C	Free throws
T1I	Free throw attempts
T1 %	Free throw percentage
RT	Total rebounds
RD	Defensive rebounds
RO	Offensive rebounds
Asis	Assists
BR	Steals
BP	Turnovers
FPC	Personal fouls
FPR	Personal fouls received
PQ	Points scored by the initial team
PI	Points scored by the centres
Val	Team evaluation
Q1	Points scored in the first quarter
Q2	Points scored in the second quarter
Q3	Points scored in the third quarter
Q4	Points scored in the fourth quarter
Q1 - Q1 (R)	Difference between points scored and received in the first quarter
Q2 - Q2 (R)	Difference between points scored and received in the second quarter
Q3 - Q3 (R)	Difference between points scored and received in the third quarter
Q4 - Q4 (R)	Difference between points scored and received in the fourth quarter

TABLE A6 Acronyms used in A3 and A4

Variable	Description
Difference	Final game difference
% RT	Rebound percentage
% RD	Defensive rebound percentage
% RO	Offensive rebound percentage
Pace	Estimated number of paces
POS	Estimated number of possessions
Seg × POS	Duration of each possession in seconds
% BP × POS	Percentage of possessions where the ball is lost
OER	Points scored per possession
OER pot	Points scored per possession ignoring lost balls
Eficacy	Points percentage
% PQ	Points percentage scored by the initial team
% PI	Points percentage scored by the centres
% T2I/TCI	Percentage of field goal attempts that are 2-point field goal attempts
BR/BP	Number of balls recovered for each lost ball
Asis/BP	Number of assists given for each lost ball
T1I/FPR	Number of free throw attempts for each personal foul received
% T1 Points	Percentage of points due to free throws
% T3 Points	Percentage of points due to 3-point throws
% Asis × Pos	Percentage of possessions where an assist is given
T2I × T3I	Number of 2-point field goal attempts for each 3-point field goal attempt
TCI × T1I	Number of field goal attempts for each free throw attempt
% BPnoF	Percentage of unforced lost balls
eFG %	Effective field goal percentage
% TS	True shooting percentage

TABLE A7 Classification of RMCF games by result and game site

Result	Visitor	Local	Total
Lose	10	2	12
Win	33	35	68
Total	43	37	80

TABLE A8 Classification of RMCF games by result and competition

Result	ACB	Cup	Euroleague	Super Cup	Total
Lose	6	0	6	0	12
Win	38	3	25	2	68
Total	44	3	31	2	80

TABLE A9 Classification of FCB games by result and game site

Result	Visitor	Local	Total
Lose	13	5	18
Win	30	31	61
Total	43	36	79

TABLE A10 Classification of FCB games by result and competition

Result	ACB	Cup	Euroleague	Super Cup	Total
Lose	10	1	6	1	18
Win	35	2	23	1	61
Total	45	3	29	2	79

TABLE A11 Classification of FCB games according to starting $Q4$ winning or losing

Result	Defeat	Win	Total
Start Q4 losing	14	7	21
Start Q4 winning	4	54	58
Total	18	61	79