Uncovering value-drivers of high performance soccer players

Maribel Serna Rodríguez[1], **Andrés Ramírez–Hassan**[2], **and Alexander Coad**[3]

Abstract

This paper tries to uncover the drivers of soccer players' market value in the five major European soccer leagues taking into account model uncertainty (variable selection) in a framework with 35 trillion potential models. For this purpose, we use a hedonic regression framework, and implement *Bayesian model averaging* (*BMA*) through *Markov chain Monte Carlo model composition* (MC^3). To deal with endogeneity issues, *instrumental variable Bayesian model averaging* (*IVBMA*) is implemented as well. We find very strong, and robust evidence, that the most important value-drivers are player's performance, participation in the national team (Senior and Under–21), age, and age–squared.

Keywords: *BMA*, *IVBMA*, MC^3, soccer, value-drivers.

JEL: Z22, Z23, L83, C11.

[1]CENTRUM Católica Graduate Business School, Pontificia Universidad Católica del Perú, Lima, Perú.
[2]School of Economics and Finance, Departament of Economics, Universidad EAFIT, Medellín, Colombia.
[3]CENTRUM Católica Graduate Business School, Pontificia Universidad Católica del Perú, Lima, Perú.

Corresponding Authors:
Maribel Serna Rodríguez, CENTRUM Católica Graduate Business School, Lima, Perú.
Email: mserna@pucp.pe

Introduction

People are the ultimate determinant of organizational performance. Traditional sources of competitive advantage related to markets, financial capital, and economies of scale have been weakened by globalization and other environmental changes (Cappelli, 1998; Pennings, Lee, & Van Witteloostuijn, 1998). Monetary compensation is one of the main instruments used by organizations to attract, motivate and retain their staff (Pfeffer, 1994). This motivates research into the role of reward and compensation systems, for individual performance (Montanari, Silvestri, & Bof, 2008).

Both theory and practice show that employees expect rewards in return for their work (DeCenzo & Robbins, 2001). An employee reward system comprises the policies, practices, and processes by which organizations reward their employees according to their contributions, skills and competencies, as well as the market value (Armstrong, 2002). In Mahoney's (1989) organization alignment theory of work payment, three compensation systems are presented according to three different pay bases: performance, work, and skills.

The debate on the relationship between different compensation systems and individual performance is central to the sports industry (Montanari et al., 2008). Sports are an example of labor-intensive contexts, where human resources strongly affect organizational performance (Keidell, 1987; Wright, Smart, & McMahan, 1995). Therefore, compensation systems can play a key role in affecting players' motivation and performance (Montanari et al., 2008).

Both the soccer players' wage and the observable variation in transfer rates can be explained, to a large extent, by the same variables (Frick, 2007). The market values of soccer players are commonly used as wage proxies (Weimar & Wicker, 2017). Soccer salaries and reported transfer prices demonstrate that players are the key component for a professional club, and warrant that their future services should be recognized as an accounting asset. Players are an asset to each soccer club, and this is how the sportsmen acquired in the transfer market are reported on the balance sheets of the clubs (Scelles, Helleu, Durand, & Bonnal, 2016).

Therefore, understanding the market value of soccer players is fundamental, as a global economic phenomenon (Ramírez Hassan & Graciano Londoño, 2017). Despite the fact that there are theoretical frameworks, such as bargaining theory and game theory (Carmichael & Thomas, 1993), to define the process of formation of a soccer player's market value; the issue is how bets to measure the implicit theoretical constructs. It seems that there is a consensus to include variables such as players' performance, age, age–squared, experience, experience–squared, and participation in national teams. However, there are many other variables that have been used that do not have such consensus (see Tables 1 and 2). This paper seeks to clarify this lack of consensus for the five major European soccer leagues.

In particular, we use a hedonic regression framework taking into account model uncertainty (variable selection), that is, we account for the possible set of models determined by the space of possible combination of regressors using *Bayesian model averaging* (*BMA*), which is implemented using *Markov chain Monte Carlo model composition* (MC^3) given the high computational burden in this application. We also perform *instrumental variable Bayesian model averaging* (*IVBMA*), which is an extension of *BMA*, that formally accounts for model uncertainty in the presence of endogeneity to check robustness regarding this issue (Eicher, Lenkoski, & Raftery, 2009).

Although BMA enjoys a long tradition in statistics (Leamer, 1978), its application in economics has only recently come into its own; being economic growth (Fernandez, Ley, & Steel, 2001b; Sala-i Martin, Doppelhofer, & Miller, 2004) and trade (Eicher, Henn, & Papageorgiou, 2012; Jetter & Ramírez Hassan, 2015) good examples of its application. Moral-Benito (2013b) provides a detailed survey on the use of BMA methods in economics.

Hedonic pricing is a method to decompose the price of a complex product into a set of implicit or hedonic prices for the individual characteristics that compose it. This technique allows to identify the observable characteristics of soccer players that provide good predictions of their future market value estimating the implicit marginal effects (Gerrard, 2001). The strength of *BMA* lies in its ability to consider model uncertainty averaging different models. In general, this approach has good statistical characteristics regarding estimation and prediction (Raftery & Zheng, 2003). *IVBMA* includes endogeneity issues in this framework (Eicher et al., 2009; Koop, León-Gonzalez, & Strachan, 2012). In particular, we follow the proposal of Karl and Lenkoski (2012). In both scenarios, exogenous and endogenous, we use an MC^3 algorithm to avoid estimating all potential models, and therefore reducing the computational burden. MC^3 draws models with the best fit but taking into account their complexity, that is, the number of regressors is a penalty term. Thus, this algorithm selects the best models using evolutionary ideas (Madigan & York, 1995).

We found 36 variables that have been proposed to predict and describe the market value (salary or transfer rate) of soccer players (see Table 2). The dependent variable is the market value of soccer players, and independent variables come from a systematic review of the literature. The latter ones are classified into two groups, according to the integrated map of market value creation and destruction of Giuliani (2013). The effect, positive or negative, on the market value of the soccer player is the benchmark for naming a variable as a value-driver or as a negative value-driver.

Model uncertainty regarding predictors is a huge concern in this setting that should be taken into account (Hansen, 2005). In particular, we use 45 independent variables in our regression exercises. This implies 35 trillion (2^{45}) possible models due to regressors' uncertainty. To the best of our knowledge, we do not find any reference that takes into account model uncertainty to analyze the value of high performance soccer players, and therefore, to uncover the most probable drivers; a fortiori, taking into account endogeneity.

In particular, our proposal performs a weighted average of the best models taking into account fit and complexity, whereas previous studies have addressed the problem presenting a single regression model. Second, we obtain the posterior inclusion probability, that is, the probability that each particular predictor has of driving the market value. Third, we present the probabilistic knowledge (posterior distribution) about the marginal effect of each potential driver taking into account model uncertainty. Fourth, we identify the most relevant predictors taking into consideration potential endogeneity problems due to feedback effects.

After this introduction, the first section contains the literature review, which shows the large number of variables determining soccer players' value. The second section preents the methodology. The third section displays the most important predictors of the soccer player market value, using the *BMA* and the *IVBMA* techniques, divided into value-drivers and negative value-drivers. Finally, the fourth section discusses the main conclusions.

Variables

Our dependent variable is the logarithm of the "market value" of soccer players taken from Transfermarkt.es that is an online community (crowd wisdom) with special interest in professional soccer. The main reason for this choice is data limitation due to compatibility between different data sources. In particular, salaries and transfer fees are not disclosed for many players, and we perform a random sampling to avoid selection bias.

The literature uses the "market values" as a proxy for player wages or transfer fees. Although there are conceptual differences between market values and transfer fees, they are comparable (He, Cachucho, & Knobbe, 2015). "Players' market values are estimates of the transfer fees that are most likely to be paid for them" (Müller, Simons, & Weinmann, 2017, pp. 612). The market value of a soccer player is an estimate of the money a soccer club would be willing to pay to make an athlete sign a contract, independent of an actual transaction (Herm, Callsen-Bracker, & Kreis, 2014). Research on judgment and decision-making supplies strong empirical and theoretical evidence that favor that statistical estimates over human (heuristic) judgments (Dawes, Faust, & Meehl, 1989), especially for complex decisions (Evans, 2006; Tversky & Kahneman, 1974) like estimating a soccer player's market value. In addition, the literature highlights the value of collective judgments, named as the "Wisdom of Crowds", for assessing the probability of future events (Surowiecki, 2005). Chen, De, Hu, and Hwang (2014), Mollick and Nanda (2015), and Brown, Rambaccussing, Reade, and Rossi (2016) provide supportive evidence of the value of this type of crowd wisdom.

In the soccer transfer market, negotiation over the price of athletes is institutionalized (Carmichael & Thomas, 1993) and it is characterized by freedom of contract players from remote countries, many buyers and sellers, and a comprehensive (albeit subjectively based) information network (Szymanski, 1997; Szymanski & Smith, 1997). For clubs involved, there are different possible agreements which can benefit them, but there is a conflict of interest over which agreement should be struck. For this reason, the eventual outcome is the result of negotiation between the clubs. As such, both bargaining theory and game theory (Nash games) provide an appropriate framework for an analysis of this market (Carmichael & Thomas, 1993). Carmichael and Thomas (1993) model the transfer fee within a Nash-bargaining framework, following Rottenberg's proposition that the amount of the fee would be the outcome of "a bargaining process"

Transfermarkt[1] website publishes monetary valuations, or "market values", for soccer players, and these are referred to regularly by researchers in sports economics as well as in the management literature (Bryson et al., 2013; Herm et al., 2014), and by sports agents during player contract

[1]The database offered in Transfermarkt and the German Kicker magazine experts, display soccer player market valuations as a substitute for undisclosed salary (Bryson, Frick, & Simmons, 2013). The reliability of these proxies is due to the next reasons: (a) the correlation between salary produced by both sources is 0.89 for a comparable sample, (b) in Kicker, valuations have been published by a stable team of experts, and in Transfermarkt, they are based on recommendations of the market value of up to 190,000 users, and c) Kicker found a correlation of 0.80 between a subsample of real wages of the Bundesliga, and their market valuations (Bryson et al., 2013).

negotiations (Peeters, 2018). This website is the leader on the soccer transfer market (Müller, Simons, & Weinmann, 2017), and Herm et al. (2014) used Brunswik's lens model to conceptualize how the Transfermarkt crowd estimates market value. In fact, Herm et al. (2014) found a strong correlation between the Transfermarkt valuations and the actual transfer fees paid, giving it a reputation of an observer whose estimations are a great reflection of the current market values. Several club officials have revealed privately that sports agents tend to refer to this website's valuations during player contract negotiations, meaning their influence on the soccer player transfer market (Peeters, 2018).

Table 1 summarizes recent empirical studies regarding value-drivers of soccer players, but at the same time illustrates a critical point: a prominent variety regarding regressors, except for a few regressors. In particular, the number of regressors fluctuates between 6 and 33, and remains unclear which variables should form part of a regression determining value of soccer players. Table 2 summarizes the 36 significant control variables from the papers in Table 1. We can see in this table that performance and age are the predominant variables in 7 out of 9 papers, followed by experience and participating in a national team, in 5 and 4 papers, respectively. Otherwise, there is a high level of diversity regarding regressors.

According to Giuliani (2013), those variables for which a positive sign is expected for their coefficients will be called value-drivers, and those variables for which a negative sign is expected for their coefficients will be called negative value-drivers. Therefore, a value-driver is a variable that generates a positive effect (improvement) on the market value of the soccer player, and a negative value-driver is a variable that produces a negative impact (deterioration) on it.

Our focus is the 2015/2016 season in the major European soccer leagues. These leagues have formed the basis of much research on soccer player salaries (Baroncelli, Lago, & Szymanski, 2004; Dobson, Goddard, & Dobson, 2001). In 2016, 14.4% of transfers (2,105 out of a total of 14,591) generated transfer fees for an amount equivalent to USD 4.79 trillion. This market is dominated again by UEFA clubs, registering a total spending on transfers of USD 3.93 trillion, and which corresponds to 82.11% of total expenditure during the year. Regarding the latter, 80.68% (corresponding to USD 3.17 trillion) comes from the Premier League (England), the Bundesliga (Germany), LaLiga (Spain), the Serie A (Italy), and the Ligue 1 (France), in their respective order, and corresponding to the five major European soccer leagues (FIFA\TMS, 2017).

We use probabilistic random stratified sampling. This sampling is justified because of the existence of populations with well-defined subsets, relative to the feature to be evaluated, indicating that there is consistency within each layer, and heterogeneity between layers. In particular, we draw 335 soccer players (67 of the Premier League, 65 of the Bundesliga, 63 of the LaLiga, 68 of the Serie A, and 72 of the Ligue 1) allowing as maximum error 0.05, and a confidence level of 95%.

We operationalize the 33 0 36 control variables in Table 2 as 48 variables, since one variable can generate other variables, e.g., the team position variable is recoded into four dummy variables such as goalkeeper, defender, midfield, and forward. Then, 43 variables, out of 48, are taken into account for multicollinearity analysis (we omit 5 squared variables which by construction generate multicollinearity). We identify 3 variables generating multicollinearity (*Bundesliga*, *Player all-star game*, and *Club world cup*). Then, we use in our estimation exercises 45 variables, which implies 35 trillion possible models.[2]

Table 1
Recent empirical studies on value drivers of soccer players

Title of paper	Authors	Dependent variable	Total of regressors	Empirical method
Economic return on schooling for soccer players	Barros (2001)	Net salary after taxes	6	Mincerian Model of OLS, Mincerian model with instrumental variable, OLS extended model, extended model with instrumental variable
A new approach to measuring player and team quality in professional team sports	Gerrard (2001)	Transfer rate	12	Hedonic price regression
Superstar effects in sport: Evidence from Italian Soccer	Lucifora and Simmons (2003)	Salary	33	OLS with fixed effects
Performance and individual characteristics as predictors of pay levels: The case of the Italian 'Serie A'	Montanari et al. (2008)	Salary	8	Standard multiple regression
Returns to stardom: Evidence from U.S. Major League Soccer	Kuethe and Motamed (2010)	Salary	15	OLS semi logarithmic model, median quantile regression
The returns to scarce talent: Footedness and player remuneration in European Soccer	Bryson et al. (2013)	Salary	18	OLS with fixed effects at the club, median quantile regression without fixed effects
Does performance consistency pay off financially for players? Evidence from the Bundesliga	Deutscher and Büschemann (2014)	Salary	14	Robust OLS, OLS with fixed effects, median quantile regression
When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community	Herm et al. (2014)	Market value	10	OLS, White´s robust standard error, quantile regression at 95%
Assortative matching using soccer data: Evidence of mobility bias	Drut and Duhautois (2015)	Ln salary	12	Pooled OLS, OLS with fixed effects

[2] Transfermarkt website is the data source (https://www.transfermarkt.es). See Appendix B, and Appendix C.

Table 2

Relationship associated with coefficient of significant variables at α=5%

Variable	Barros (2001)	Gerrard (2001)	Lucifora and Simmons (2003)	Montanari et al. (2008)	Kuethe and Motamed (2010)	Bryson et al. (2013)	Deutscher and Büschemann (2014)	Herm et al. (2014)	Drut and Duhautois (2015)
Performance	+	+	+	+		+	+		+
Age		+	+		-	+	+	-	+
Age squared		-	-		+	-	-		-
National team					+	+	+		+
Team position						+	+		+
Goals		+						+	
Goals squared			+						
Experience	+	+		+	+				+
Experience squared	-	-			-				-
Assist								+	
Nationality					+	+			+
Goals previous season		+	+			+			
League									
Matches			+				+		
Matches squared			-				-		
Legs						+			
Assistances matches						+			
Team performance				+					+
Club administration								+	
Sports agent									
Height						-			
Height squared						+			
Substitute						+			
Public attention								+	
Member union association	+								
International players team									+
Superstar 1			+						
Superstar 2			+						
Player all-star game									
Designated player					+				
Under-21 but not senior			+						
Appearances UEFA champions						+			
Recognition under-21		+							
International recognition		+							
Sportsman reputation				+					
Team Experience				+					

Methodology

BMA takes into account model uncertainty performing a weighted average of the estimates over the set of potential models $\mathcal{M} = \{M_1, M_2, \ldots, M_{2^k}\}$, such that k explanatory variables implies 2^k potential models.[3] Therefore, we use Bayes' rule to calculate the posterior model probability $p(M_r|\boldsymbol{y})$, that is, the probability that a particular configuration of regressors generates the data set that we actually observe. This probability is the weight associated with model M_r in the weighted average.

[3]For instance, 3 regressors (x_1, x_2 and x_3) would imply 8 possible models (2^3): $\mathcal{M} = \{\{\emptyset\}, \{x_1\}, \{x_2\}, , \{x_3\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}\}$. All models include intercept, then the first model is just with intercept.

$$p(M_r|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|M_r)p(M_r)}{\sum_{j=1}^{2^k} p(\boldsymbol{y}|M_j)p(M_j)}, \tag{1}$$

where $p(M_r)$ is the prior model probability, that is, the probability that we assign to a particular model prior to observe the data set,[4] and $p(\boldsymbol{y}|M_r)$ is the marginal likelihood, which is interpreted as the weight of evidence for the model M_r. Unlike the maximized likelihood, the marginal likelihood has an implicit penalty for model complexity; asymptotically its logarithm converges to $-\frac{BIC}{2}$, where BIC is the Bayesian information criterion. See Appendix A for details.

Under BMA, the expected value of marginal effects is given by the weighted average taking into account the set of potential models, that is,

$$E(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{r=1}^{2^k} p(M_r|\boldsymbol{y})E(\boldsymbol{\theta}|\boldsymbol{y}, M_r),$$

where $E(\boldsymbol{\theta}|\boldsymbol{y}, M_r) = \int_{\boldsymbol{\Theta}_r} \boldsymbol{\theta}_r p(\boldsymbol{\theta}_r|\boldsymbol{y}, M_r) d\boldsymbol{\theta}_r$ is the expected value of $\boldsymbol{\theta}$ under model M_r given the data set. $p(\boldsymbol{\theta}_r|\boldsymbol{y}, M_r)$ is the posterior distribution of $\boldsymbol{\theta}_r$ under M_r (see equation 3).

Similarly the variance is

$$Var(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{r=1}^{2^k} p(M_r|\boldsymbol{y})Var(\boldsymbol{\theta}|\boldsymbol{y}, M_r) + \sum_{r=1}^{2^k} p(M_r|\boldsymbol{y})(E(\boldsymbol{\theta}|\boldsymbol{y}) - E(\boldsymbol{\theta}|\boldsymbol{y}, M_r))^2, \tag{2}$$

where $Var(\boldsymbol{\theta}|\boldsymbol{y}, M_r)$ is the variance of $\boldsymbol{\theta}$ under model M_r given the data set.

Equation 2 highlights how BMA accounts for model uncertainty. The first term is the weighted variance for each model, averaged over all potential models, and the second term indicates how stable the marginal effects are across models. The more the marginal effects differ between models, the greater is the posterior variance.

In our setting each potential model has the form $\boldsymbol{y} = \beta_0 \boldsymbol{i}_N + \boldsymbol{X}_r \boldsymbol{\beta}_r + \boldsymbol{\epsilon}_r$ where $\boldsymbol{y}$ is the logarithm of the soccer player market value, $\boldsymbol{i}_N$ is a N-dimensional vector of 1's, $\boldsymbol{X}_r$ is a $N \times k_r$ matrix of a particular combination of soccer player characteristics that could influence their market value, β_0 and $\boldsymbol{\beta}_r$ are location parameters, and $\boldsymbol{\epsilon}_r \sim \mathcal{N}(\boldsymbol{0}_N, h^{-1}\boldsymbol{I}_N)$ is a N-dimensional vector of

[4] It is intuitive to assign equal prior probability to each model, that is, $p(M_r) = \frac{1}{2^k}$. However, this choice gives more prior probability to the set of models of medium size (think about the k-th row of Pascal's triangle). So, we use the Beta-Binomial prior proposed by Ley and Steel (2009).

stochastic errors, $h = 1/\sigma^2$ is the precision parameter.

We can use Bayes' rule again to obtain the posterior distribution of the set of parameters, $\boldsymbol{\theta}_r = [\boldsymbol{\beta}_r', h, \beta_0]'$. This distribution is an updated version of the probabilistic knowledge that we have about $\boldsymbol{\theta}_r$ given the data set.

$$p(\boldsymbol{\theta}_r|\boldsymbol{y}, M_r) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}_r, M_r)p(\boldsymbol{\theta}_r|M_r)}{p(\boldsymbol{y}|M_r)}, \tag{3}$$

where $p(\boldsymbol{\theta}_r|M_r)$ is the probability that we assign to the parameters prior to observe the data set given model M_r, and $p(\boldsymbol{y}|\boldsymbol{\theta}_r, M_r)$ is the likelihood function given model M_r.

We use "non–informative" priors for the common parameters to all models, that is, β_0 and h. The purpose is to try to be as "objective" as possible in our estimation exercise. In particular, $p(h) \propto 1/h$ and $p(\beta_0) \propto 1$. In addition, $\boldsymbol{\beta}_r|h \sim \mathcal{N}(\mathbf{0}_r, h^{-1}[g_{k_r}\boldsymbol{X}_r'\boldsymbol{X}_r]^{-1})$, that is, we assume a priori that the regressors do not have an effect on the soccer player market value, and the covariance matrix is guided by sample information, such that its precision depends on the number of observations and the number of regressors in each model. Accordingly, Fernandez, Ley, and Steel (2001a) suggest selecting $g_{k_r} = 1/k_r^2$ if $N \leq k_r^2$ or $g_{k_r} = 1/N$ if $N > k_r^2$. We show in Appendix A the resulting posterior distribution.

BMA uses equations 1 and 3 to take into account model uncertainty. In particular, the posterior probabilistic knowledge about the parameters is given by

$$p(\boldsymbol{\theta}_r|\boldsymbol{y}) = \sum_{r=1}^{2^k} p(\boldsymbol{\theta}_r|\boldsymbol{y}, M_r)p(M_r|\boldsymbol{y}). \tag{4}$$

This is a weighted average of the posterior distributions over the set of potential models. This approach is superior to any single selected model (Raftery, Madigan, & Hoeting, 1997a).

Despite the beauty of this probabilistic approach, it can be impractical if the set of models is huge. In our specific case, there are 45 potential drivers, this implies 35 trillion potential models. Therefore, we use Markov chain Monte Carlo model composition (MC^3) (Madigan, York, & Allard, 1995). The main objective of this algorithm is to avoid calculating the posterior model probability of all potential models, and their parameters posterior distributions, and just to focus on the set of models with the highest posterior probabilities. Thus, models whose posterior probability is close to 0 are eliminated from equation 4, as they do not contribute significantly. Raftery, Madigan, and

Hoeting (1997b) found that in linear regression settings as ours, less than 25 models contribute significantly.

MC^3 is based on evolutionary ideas, where an actual model competes against a candidate model, and the "best-fit" model survives. Even in the case that the candidate model is "less-fit" than the actual model, the former has a probability to survive. This is done to avoid being trapped in local solutions. The "fit" criterion is a mix between model fit and model complexity (number of regressors). Thus, MC^3 generates a sequence of models whose posterior probabilities are the highest among all models after S iterations ($S << 2^k$). See Appendix A for technical details.

Once we perform MC^3, we can calculate the posterior inclusion probability of variable $\boldsymbol{x}_l$, $l = 1, 2, \dots, k$, that is, the probability that regressor $\boldsymbol{x}_l$ is associated with market value.

$$PIP(\boldsymbol{x}_l) = \sum_{r=1}^{2^k} p(M_r|\boldsymbol{y}) \times I_{l,r}, \tag{5}$$

where $I_{l,r} = \left\{ \begin{array}{lll} 1 & if & \boldsymbol{x}_l \in M_r \\ & & \\ 0 & if & \boldsymbol{x}_l \notin M_r \end{array} \right\}$.

Kass and Raftery (1995) suggest that $PIP < 0.5$ is evidence against the effect, $0.5 \leq PIP < 0.75$ is weak evidence for the effect, $0.75 \leq PIP < 0.95$ is positive evidence for the effect, $0.95 \leq PIP < 0.99$ is strong evidence, and $PIP \geq 0.99$ is very strong evidence.

Endogeneity can be a particular concern in our application (Deutscher & Büschemann, 2014). Therefore, we address this issue using *IVBMA* (Karl & Lenkoski, 2012). This technique resembles the process of *BMA*, but using a Bayesian two-stage procedure based on conditional Bayes factors (Dickey & Gunel, 1978), similar in spirit as two-stage least squares. Therefore, we take into account model uncertainty and endogeneity simultaneously. See Appendix A for details.

Results

Following the methodological outline described in the previous section, first the variance inflation factor (VIF) procedure was applied and next the *BMA* methodology was implemented.[5] Table 3 shows 40 of 45 variables to take into consideration for applying *BMA*. The five additional

[5]All our calculations were performed in R software (R Core Team, 2018).

variables to be incorporated in the *BMA* analysis are *Age squared*, *Goals squared*, *Experience squared*, *Matches squared*, and *Height squared*.

Table 3
*Final variables according to VIF results**TAMAÑO COLUMNAS TABLA*

Variables				
Performance	Age	National team	Defender	Midfield
Forward	Goals	Experience	Assists	Concacaf
Conmebol	Afc	UEFA	Goals previous season	Premier league
Serie A	Ligue 1	Matches	Ambidextrous	Right leg
Assistances matches	First division	Second division	Club administration	Sport agent
Height	Substitute	Public attention*	Member union association	International players team
Euro UEFA	UEFA champion	UEFA European	UEFA super cup	Team experience
Superstar 2	Under-21 but no senior	Appearances UEFA Champions	Recognition under 21	Fifa cup

*We use natural logarithm of this variable.

We present in Table 4 Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS) estimates using all variables for comparison purposes. 2SLS and *IVBMA* try to address the potential endogeneity issue between *Market value* and *Performance* (Deutscher & Büschemann, 2014).[6] In both cases, we use *Coach* and *Change of team* as instruments. The intuition is that our measure of performance (player appearances last season, 2015–2016) is associated with these two variables; coaches make decisions about soccer players' appearances, and change of team affects players' appearances as well. We expect that these instruments do not have a direct effect on market value once we control for other characteristics in the main equation. The p-values of the Sargan test (overidentification restrictions) are 0.11 and 0.08 in the case of 2SLS (Sargan, 1958) and *IVBMA* (Lenkoski, Eicher, & Raftery, 2014), respectively. This means no rejection of the null hypothesis of no correlation between the instruments and the stochastic error in the main equation.[7]

The number of iterations is 1,200,000, and the number of burn-in draws for the *MCMC* sampler is 200,000. As a result, the algorithm visited 109,483 out of 2^{45} potential models, the correlation between the analytic and *MCMC* posterior model probabilities is 0.99, that is a near perfect correlation that suggests a good performance of the algorithm (Fernandez, Ley, & Steel, 2001a).

[6]The p-value of the endogeneity test using the J statistic proposed by Davidson and MacKinnon (1981) correcting for heteroscedasticity is 0.089. The 95% credible interval for the covariance between *Market value* and *Performance* using IVBMA is $(-4.76, -1.26)$ (see Appendix A for technical details). Both criteria suggest endogeneity.

[7]An underidentification test (Anderson, 1951) rejects the null hypothesis of correlation between the instruments and the stochastic error in the main equation, with p-value 0.00. A Weak identification test (Cragg & Donald, 1993) has a F statistic equal to 4.92, this means a p-value equal to 0.00 which implies rejection of the null hypothesis of weak instruments. *PIPs* of *Coach* and *Change of team* are 0.29 and 0.90 in the first stage of *IVBMA*. Results are available upon request.

Table 4
OLS and 2SLS results

Variable	OLS Coefficients*	Standard error OLS	2SLS Coefficients	Standard error 2SLS
Performance	2.202e-02*	(0.005)	7.568e-02*	(0.032)
Age	7.528e-01*	(0.195)	6.940e-01*	(0.180)
Age squared	-1.51e-02*	(0.004)	-1.39e-02*	(0.003)
National team	6.775e-01*	(0.108)	5.866e-01*	(0.130)
Defender	1.811e-01	(0.220)	9.091e-02	(0.195)
Midfield	1.138e-01	(0.246)	6.121e-02	(0.215)
Forward	1.405e-01	(0.287)	1.668e-01	(0.255)
Goals	1.270e-02*	(0.004)	1.510e-02*	(0.005)
Goals squared	-2.90e-05*	(0.000)	-1.89e-05	(0.000)
Experience	5.725e-02	(0.064)	5.790e-02	(0.060)
Experience squared	-3.57e-03	(0.002)	-3.44e-03	(0.002)
Assists	-1.66e-02	(0.016)	-7.49e-02	(0.042)
Concacaf	-3.12e-01	(0.245)	-4.60e-01	(0.470)
Conmebol	3.201e-01	(0.209)	3.306e-01	(0.216)
Afc	-4.86e-01	(0.287)	-5.53e-01	(0.440)
UEFA	2.476e-01	(0.168)	6.353e-02	(0.212)
Goals previous season	1.432e-02	(0.029)	-1.01e-01	(0.078)
Premier league	4.923e-01*	(0.201)	5.544e-01*	(0.197)
SerieA	-3.23e-04	(0.171)	-7.03e-02	(0.186)
Ligue 1	-2.41e-01	(0.152)	-1.77e-01	(0.186)
Matches	7.139e-03*	(0.002)	-1.25e-03	(0.005)
Matches squared	-1.79e-05*	(0.000)	-1.77e-06	(0.000)
Ambidextrous	-7.90e-01*	(0.207)	-1.0e+00*	(0.372)
Right leg	-6.59e-02	(0.116)	-6.80e-02	(0.117)
Assistances matches	6.850e-07*	(0.000)	7.290e-07*	(0.000)
First division	6.243e-01*	(0.165)	7.544e-01*	(0.195)
Second division	-1.03e-02	(0.064)	1.151e-01	(0.099)
Club administration	5.425e-04	(0.001)	1.031e-03	(0.001)
Sport agent	1.693e-01	(0.122)	1.437e-01	(0.126)
Height	-1.46e+01	(30.29)	2.30e+00	(36.19)
Height squared	4.23e+00	(8.402)	-5.08e-01	(9.954)
Substitute	-2.42e-02*	(0.008)	-2.48e-02*	(0.009)
Public attention	2.350e-02	(0.033)	1.496e-02	(0.032)
Member union association	4.274e-01*	(0.177)	4.544e-01*	(0.192)
International players team	-1.90e-03	(0.294)	-4.29e-01	(0.423)
Superstar 2	6.647e-01	(0.498)	1.11e+00	(1.055)
Under-21 but not senior	5.251e-01	(0.308)	2.155e-01	(0.457)
Appearances UEFA Champions	4.133e-02	(0.224)	2.819e-02	(0.204)
Recognition under-21	3.213e-01*	(0.113)	4.035e-01*	(0.129)
Fifa cup	6.886e-01	(0.415)	9.436e-01	(0.969)
Euro UEFA	-1.2e+00*	(0.223)	-9.90e-01	(0.910)
UEFA champion	3.404e-01	(0.528)	-8.34e-01	(1.253)
UEFA European	4.693e-01	(0.390)	7.423e-01	(0.577)
UEFA super cup	2.826e-01	(0.385)	-5.97e-01	(1.068)
Team experience	6.703e-02*	(0.030)	6.017e-02	(0.036)
Constant	1.49e+01	(26.90)	8.097e-01	(32.68)

* Statistically significant at 5%.

According to the guidelines of Kass and Raftery (1995), there is very strong evidence regarding six variables determining soccer player market value, $PIP \geq 0.99$, which mean that practically all the best models include these variables. Five of these variables are value-drivers, and one of them is a negative value-driver. The value-drivers of soccer players' market value are player appearances in the last season ($Performance$), participation of soccer player in a Senior national team (*National team*), age of athlete (*Age*), goals scored during his career (*Goals*), participation of a soccer player in the Under–21 national team ($Recognition\ under - 21$). The negative value-driver is the square of the age (*Age squared*). In addition, there is positive evidence, $0.75 \leq PIP < 0.95$, that team's performance (*First division*, $PIP = 0.90$) and team's market value, excluding player market value (*Club administration*, $PIP = 0.86$) are associated with soccer player market value. After these eight variables, the PIP shows a remarkable decrease (see Table 5).

We observe comparing OLS and *BMA* results that OLS suggests 17 statistically significant variables at 5% (see Table 4), whereas *BMA* suggests just six (or eight in the best scenario) (see Table 5). This is because OLS does not take into account model uncertainty, therefore it overestimates its precision.

Figure 1 displays the posterior model size distribution and provides evidence concerning the eight variables that define the soccer player market value. Table 6 confirms that the posterior model probability (PMP) of taking into consideration the eight most relevant variables is 10.92%. Compared to the first best model, the PMP for the second best model is 5.78%, that is obtained after adding the variable *Goals squared*.

The posterior mean (Post Mean), column 3 in Table 5, displays the coefficients averaged over all models. The density functions of the most relevant coefficients are displayed in Figure 2. The marginal density for *Age squared* shows that the expected value of this negative value-driver is -0.0197. At this point, we remark that the probability density functions for five variables are unimodal, and the probability density function for the variable *Goals* is multimodal with modes at 0.007 and 0.015.

In addition, Table 5 shows the posterior standard deviations (Post SD, column 3), and the conditional posterior sign (Cond.Pos.Sign, column 4); a value of 1 means that the control variable has a positive effect on soccer player market value in all models. On the other hand, a value equal to 0 indicates that the control variable has a negative effect on soccer player market value in all models

Table 5

BMA results

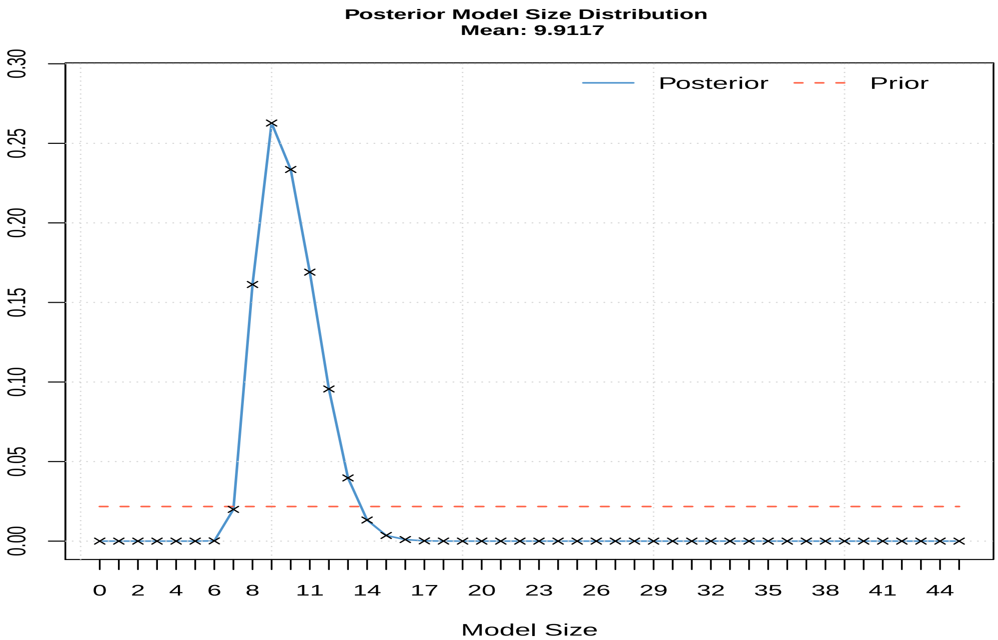
Variable	PIP	Post Mean	Post SD	Cond.Pos.Sign*
Performance	1.000	2.994e-02	5.410e-03	1.000
National team	1.000	6.259e-01	1.051e-01	1.000
Age	1.000	1.01e+00	1.331e-01	1.000
Age squared	1.000	-1.97e-02	2.454e-03	0.000
Goals	0.993	9.502e-03	3.989e-03	1.000
Recognition under-21	0.992	4.563e-01	1.094e-01	1.000
First division	0.905	5.617e-01	2.432e-01	1.000
Club administration	0.862	1.945e-03	9.523e-04	1.000
Substitute	0.358	-8.77e-03	1.273e-02	0.000
Team experience	0.294	2.657e-02	4.456e-02	1.000
Goals squared	0.290	-1.31e-05	2.225e-05	0.002
Matches	0.196	1.451e-03	3.180e-03	1.000
Matches squared	0.182	-3.46e-06	7.802e-06	0.013
Assistances matches	0.141	1.021e-07	2.693e-07	1.000
Ligue 1	0.119	-3.87e-02	1.150e-01	0.000
Conmebol	0.112	3.715e-02	1.163e-01	1.000
Public attention	0.084	5.465e-03	1.989e-02	1.000
Premier league	0.077	2.396e-02	9.387e-02	1.000
Ambidextrous	0.036	-1.98e-02	1.185e-01	0.000
Afc	0.025	-1.54e-02	1.149e-01	0.000
Under-21 but not senior	0.022	1.275e-02	1.037e-01	1.000
Forward	0.017	3.761e-03	3.776e-02	0.987
Appearances UEFA Champions	0.017	4.316e-03	4.195e-02	1.000
Fifa cup	0.016	1.880e-02	1.862e-01	1.000
Member union association	0.016	3.385e-03	4.116e-02	0.698
Sport agent	0.016	2.636e-03	2.550e-02	1.000
UEFA	0.015	3.020e-03	3.794e-02	0.702
UEFA European	0.014	8.161e-03	9.065e-02	1.000
Concacaf	0.012	-4.93e-03	6.166e-02	0.000
Euro UEFA	0.011	-9.41e-03	1.265e-01	0.000
Superstar 2	0.010	8.057e-03	1.198e-01	1.000
Experience squared	0.010	-1.09e-05	1.825e-04	0.003
Height squared	0.010	1.068e-03	8.938e-02	0.927
Height	0.010	5.030e-03	3.261e-01	0.928
Serie A	0.009	6.988e-04	1.596e-02	0.816
Defender	0.009	7.819e-04	1.345e-02	0.989
Goals previous season	0.009	1.974e-04	3.621e-03	1.000
Second division	0.009	4.329e-04	7.542e-03	0.981
Experience	0.009	4.399e-05	3.625e-03	0.465
Midfield	0.009	-3.85e-04	1.091e-02	0.085
International players team	0.008	8.391e-04	3.037e-02	0.822
Assists	0.008	-7.64e-05	2.134e-03	0.163
UEFA champion	0.008	-2.39e-03	8.442e-02	0.109
UEFA super cup	0.008	-1.32e-03	7.919e-02	0.158
Right leg	0.007	-2.15e-04	9.435e-03	0.177

* Cond.Pos.Sign is the conditional posterior sign. It is the posterior fraction of times the sign of the variable is positive.

Table 6
Variables included in the top 10 models

Variables	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
Performance	1	1	1	1	1	1	1	1	1	1
National team	1	1	1	1	1	1	1	1	1	1
Age	1	1	1	1	1	1	1	1	1	1
Age squared	1	1	1	1	1	1	1	1	1	1
First division	1	1	1	1	1	1	1	1	0	1
Club administration	1	1	1	1	1	1	0	1	1	1
Defender	0	0	0	0	0	0	0	0	0	0
Midfield	0	0	0	0	0	0	0	0	0	0
Forward	0	0	0	0	0	0	0	0	0	0
Goals	1	1	1	1	1	1	1	1	1	1
Goals squared	0	1	0	0	1	0	0	0	0	1
Experience	0	0	0	0	0	0	0	0	0	0
Experience squared	0	0	0	0	0	0	0	0	0	0
Assists	0	0	0	0	0	0	0	0	0	0
Concacaf	0	0	0	0	0	0	0	0	0	0
Conmebol	0	0	0	0	0	0	0	0	0	0
Afc	0	0	0	0	0	0	0	0	0	0
UEFA	0	0	0	0	0	0	0	0	0	0
Goals previous season	0	0	0	0	0	0	0	0	0	0
Premier league	0	0	0	0	0	0	0	0	0	0
Serie A	0	0	0	0	0	0	0	0	0	0
Ligue 1	0	0	0	0	0	0	0	0	0	0
Matches	0	0	0	0	0	0	0	1	0	0
Matches squared	0	0	0	0	0	0	0	1	0	0
Ambidextrous	0	0	0	0	0	0	0	0	0	0
Right leg	0	0	0	0	0	0	0	0	0	0
Assistances matches	0	0	0	0	0	0	1	0	0	0
Second division	0	0	0	0	0	0	0	0	0	0
Sport agent	0	0	0	0	0	0	0	0	0	0
Height	0	0	0	0	0	0	0	0	0	0
Height squared	0	0	0	0	0	0	0	0	0	0
Substitute	0	0	1	1	0	0	0	1	0	1
Member union association	0	0	0	0	0	0	0	0	0	0
International players team	0	0	0	0	0	0	0	0	0	0
Superstar 2	0	0	0	0	0	0	0	0	0	0
Under 21 but not senior	0	0	0	0	0	0	0	0	0	0
Appearances UEFA Champions	0	0	0	0	0	0	0	0	0	0
Recognition under 21	1	1	1	1	1	1	1	1	1	1
Fifa cup	0	0	0	0	0	0	0	0	0	0
Euro UEFA	0	0	0	0	0	0	0	0	0	0
UEFA champion	0	0	0	0	0	0	0	0	0	0
UEFA European	0	0	0	0	0	0	0	0	0	0
UEFA super cup	0	0	0	0	0	0	0	0	0	0
Team experience	0	0	0	1	1	1	0	0	0	1
Public attention	0	0	0	0	0	0	0	0	0	0
PMP (Exact)	0.109	0.058	0.045	0.032	0.026	0.025	0.023	0.020	0.018	0.017
PMP (MCMC)	0.106	0.060	0.044	0.033	0.026	0.025	0.022	0.022	0.018	0.018

Figure 1. BMA: Prior and posterior model size

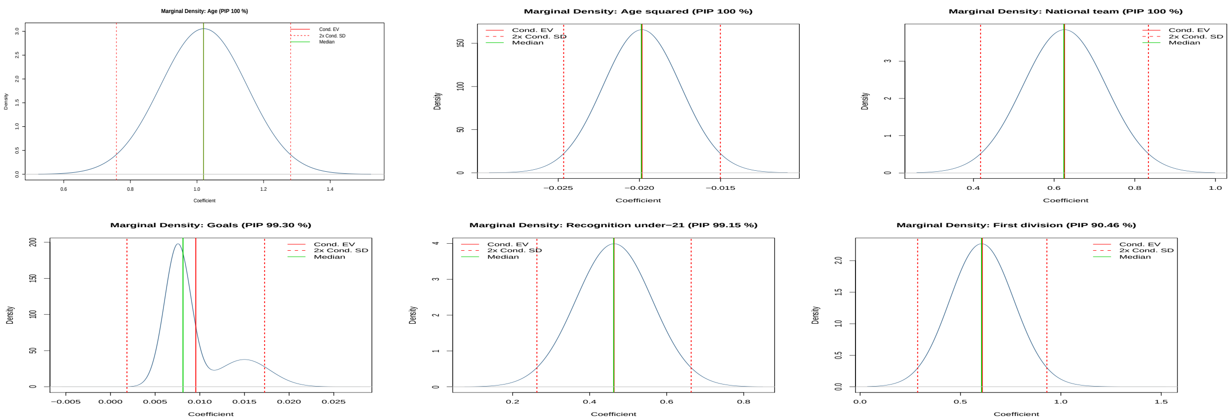


(Zeugner, 2011). Therefore, the variables *Age*, *Performance*, *National team*, *Goals*, *Recognition under* $-$ 21, *First division*, and *Club administration* increase the soccer player market value in all relevant models.

The signs of these variables' coefficients are consistent with previous analyses. Particularly, the same results have been obtained for the value-drivers such as *Performance* (Barros, 2001; Bryson et al., 2013; Deutscher & Büschemann, 2014; Drut & Duhautois, 2015; Gerrard, 2001; Lucifora & Simmons, 2003; Montanari et al., 2008), *National team* (Bryson et al., 2013; Deutscher & Büschemann, 2014; Drut & Duhautois, 2015; Kuethe & Motamed, 2010), *Age* (Bryson et al., 2013; Deutscher & Büschemann, 2014; Drut & Duhautois, 2015; Gerrard, 2001; Lucifora & Simmons, 2003), *Goals* (Gerrard, 2001; Herm et al., 2014), and *Recognition under* $-$ 21 (Gerrard, 2001). Regarding the expected sign for the negative value-driver *Age squared*, equivalent outcomes have been reached by Drut and Duhautois (2015), Deutscher and Büschemann (2014), Bryson et al. (2013), Lucifora and Simmons (2003), and Gerrard (2001).

Table 7 displays the second-stage results from using the *IVBMA* mechanism in order to take into account potential endogeneity between *Market value* and *Performance* (Deutscher & Büschemann, 2014). Results are very similar to *BMA*, except that *First division* has a $PIP = 1$,

Figure 2. BMA: Posterior probability density of the most important coefficients



and there is no very strong evidence for *Goals* ($PIP = 0.64$). In addition, there is positive evidence for *Premier league* ($PIP = 0.80$).

We found that 2SLS identifies 12 statistically significant variables at 5% (see Table 4), whereas *IVBMA* identifies just 6 (7 taking into account *Premier league*). Again, we have that 2SLS over-estimates its precision by not considering model uncertainty.

Conclusions

This paper seeks to provide evidence about which variables are more likely to determine the soccer player market value in the five major European soccer leagues, taking into account model uncertainty (variable selection). With a variety of potential drivers suggested in the empirical literature, and the lack of a comprehensive theoretical framework that defines specifically how to measure the implicit theoretical constructs; determining the most probable drivers associated with soccer market value becomes difficult. However, *BMA* allows us to take into account model uncertainty based on a strong probability theory framework. In addition, *IVBMA* allows us to control for potential endogeneity issues between performance and market value.

The sample considers 45 potential determinants of the soccer player market value, meaning 35 trillion possible models. The results from *BMA* estimations suggest that the prediction of soccer player market value is given by six variables, five of them are value-drivers and one of them is a negative value-driver. These variables are player appearances, participation of soccer player in the Senior national team, age of the athlete, square of the age, goals scored during his career, and participation of soccer player in the Under-21 national team. These results are robust when

Table 7

IVBMA results

Variable	PIP	Mean	SD	2.5 %	50 %	97.5 %
Performance	1.000	5.401e-02	9.020e-03	0.039	0.053	0.076
National team	1.000	6.794e-01	1.130e-01	0.459	0.678	0.903
Age	1.000	9.957e-01	7.416e-02	0.832	1.005	1.126
Age squared	1.000	-1.97e-02	1.415e-03	-0.02	-0.02	-0.02
First division	1.000	8.364e-01	1.846e-01	0.472	0.836	1.186
Recognition under-21	0.997	4.556e-01	1.113e-01	0.238	0.456	0.668
Premier league	0.799	3.046e-01	2.065e-01	0.000	0.333	0.674
Ligue 1	0.692	-2.48e-01	2.096e-01	-0.63	-0.27	0.000
Conmebol	0.665	2.624e-01	2.455e-01	0.000	0.261	0.762
Ambidextrous	0.658	-3.84e-01	3.692e-01	-1.11	-0.38	0.000
Fifa cup	0.644	6.084e-01	6.933e-01	-0.22	0.453	2.101
Goals	0.638	4.455e-03	3.586e-03	0.000	0.006	0.010
Appearances UEFA Champions	0.622	2.302e-01	2.236e-01	0.000	0.227	0.663
Constant	0.497	1.720e-01	6.546e-01	-1.16	0.000	1.747
Forward	0.456	1.438e-01	1.947e-01	0.000	0.000	0.561
UEFA champion	0.453	1.914e-01	5.048e-01	-0.68	0.000	1.495
UEFA European	0.448	2.185e-01	3.907e-01	-0.23	0.000	1.217
Euro UEFA	0.441	-1.82e-01	4.811e-01	-1.45	0.000	0.614
Height	0.425	1.599e-01	4.492e-01	-0.57	0.000	1.339
Afc	0.418	-1.78e-01	3.284e-01	-1.01	0.000	0.181
UEFA super cup	0.413	1.240e-01	4.564e-01	-0.74	0.000	1.369
Superstar 2	0.409	5.784e-02	4.499e-01	-0.92	0.000	1.235
Under-21 but not senior	0.398	1.601e-01	3.071e-01	-0.16	0.000	0.987
Concacaf	0.304	-6.52e-02	2.306e-01	-0.73	0.000	0.310
UEFA	0.296	6.780e-02	1.422e-01	-0.02	0.000	0.468
Sport agent	0.288	5.195e-02	1.048e-01	0.000	0.000	0.349
Substitute	0.273	-6.52e-03	1.142e-02	-0.03	0.000	0.000
International players team	0.248	2.563e-02	1.609e-01	-0.30	0.000	0.498
Goals previous season	0.232	-1.93e-02	4.064e-02	-0.14	0.000	0.000
Member union association	0.228	3.875e-02	1.099e-01	-0.06	0.000	0.376
Height squared	0.227	2.111e-02	1.372e-01	-0.23	0.000	0.411
Team experience	0.191	1.224e-02	2.880e-02	0.000	0.000	0.100
Public attention	0.174	1.074e-02	2.645e-02	0.000	0.000	0.093
Serie A	0.145	2.609e-03	6.347e-02	-0.14	0.000	0.171
Defender	0.140	1.531e-02	6.237e-02	-0.02	0.000	0.221
Midfield	0.125	4.433e-03	5.802e-02	-0.09	0.000	0.154
Right leg	0.108	-5.85e-03	4.038e-02	-0.13	0.000	0.034
Club administration	0.083	1.544e-04	5.293e-04	0.000	0.000	0.002
Second division	0.081	3.554e-03	2.295e-02	0.000	0.000	0.077
Experience	0.022	1.235e-04	3.560e-03	0.000	0.000	0.000
Matches	0.002	-6.73e-07	7.960e-05	0.000	0.000	0.000
Experience squared	0.001	-9.59e-08	1.822e-05	0.000	0.000	0.000
Assistances matches	0.000	0.00e+00	0.00e+00	0.000	0.000	0.000
Goals squared	0.000	0.00e+00	0.00e+00	0.000	0.000	0.000
Matches squared	0.000	0.00e+00	0.00e+00	0.000	0.000	0.000

controlling for endogeneity issues between value market and performance, and seems to confirm the consensus in the literature.

These outcomes seem to highlight some interesting aspects of the high level soccer players market. First, participation in national teams, and in particular the Under-21 team, is a promising signal regarding future market value, even among the best soccer players in the world. And second, soccer players maximize their market value at approximately 26 years old. SEGÚN DATOS, TABLA 5 Y NO 7. CON DATOS DE TABLA 7 SERÍAN 25 AÑOS[8]

Although our approach has good statistical characteristics, we should name some of its limitations. First, our approach is cross-sectional, this implies that it does not take into account dynamics and unobserved heterogeneity. Future research may consider these aspects. For instance, León-González and Montolio (2015); Moral-Benito (2013a) propose *BMA* in panel data settings. In addition, our statistical framework also demands many regressors to test which subsets are more compelling. This may impose some restrictions when extending our application to panel data settings. Another implicit assumption in BMA is that it assumes that the data generating process (DGP) is one in the set of potential models. This may not be true. However, BMA minimizes the Kullback-Leibler divergence (relative entropy) to the ($d.g.p$).[9] Moreover, MC^3 is a greedy algorithm. In particular, it is an algorithmic paradigm based on binary decisions (local optimum) at each stage with the objective of finding a global optimum. In many problems, it can be possible that MC^3 does not identify the (DGP), but nonetheless it may identify models that are "close" to the DGP in a reasonable amount of time.

There is great potential for more research focused on understanding the market value of athletes in different sports like basketball, baseball, and cycling among others. In fact, soccer players could be the target again if the analysis focuses on specific leagues and positions. This would be the opportunity to research what are the value-drivers and the negative value-drivers for different leagues and sports.

[8] $\widehat{Exp}^* = -1.0134/(-2 \times -0.0197)$, see Table 5.

[9] The Kullback-Leibler divergence can be interpreted as the amount of information that is lost when *BMA* is used rather than *DGP*.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, *22*, 327–351.

Armstrong, M. (2002). *Employee reward.* CIPD Publishing.

Baroncelli, A., Lago, U., & Szymanski, S. (2004). Il business del calcio. *Egea, Milano.*

Barros, C. P. (2001). Economic return on schooling for soccer players. *Journal of Sports Economics*, *2*(4), 369–378.

Brown, A., Rambaccussing, D., Reade, J. J., & Rossi, G. (2016). Using social media to identify market inefficiencies: evidence from twitter and betfair.

Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: footedness and player remuneration in european soccer. *Journal of Sports Economics*, *14*(6), 606–628.

Cappelli, P. (1998). *New deal at work.* Harvard Business School Press.

Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: theory and evidence. *Applied Economics*, *25*(12), 1467–1476. Retrieved from `https://doi.org/10.1080/00036849300000150`

Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367–1403.

Cragg, J., & Donald, S. (1993). Testing identifiability and specification in instrumental variables models. *Econometric Theory*, *9*, 222–240.

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, *49*, 781–793.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674.

DeCenzo, D. A., & Robbins, S. P. (2001). *Human resource management* (7th ed.). Wiley, NJ.

Deutscher, C., & Büschemann, A. (2014). Does performance consistency pay off financially for players? evidence from the bundesliga. *Journal of Sports Economics*, *17*(1), 27–43.

Dickey, J. M., & Gunel, E. (1978). Bayes factors from mixed probabilities. *Journal of the Royal Statistical Society. Series B*, *40*(1), 43–46.

Dobson, S., Goddard, J. A., & Dobson, S. (2001). *The economics of football.* Cambridge University Press Cambridge.

Drut, B., & Duhautois, R. (2015). Assortative matching using soccer data: Evidence of mobility bias. *Journal of Sports Economics*, *18*(5), 1–17.

Eicher, T. S., Henn, C., & Papageorgiou, C. (2012). Trade creation and diversion revisited: Accounting for model uncertainty and natural trading partner effects. *Journal of Applied Econometrics*, *27*(2), 296–321.

Eicher, T. S., Lenkoski, A., & Raftery, A. E. (2009). Bayesian model averaging and endogeneity under model uncertainty: an application to development determinants. *Center for Statistics and the Social Sciences University of Washington, Working Paper no*, *94*.

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, *13*(3), 378–395.

Fernandez, C., Ley, E., & Steel, M. F. (2001a). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, *100*(2), 381–427.

Fernandez, C., Ley, E., & Steel, M. F. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, *16*(5), 563–576.

FIFA\TMS. (2017). *Global transfer market report 2017.* Retrieved from `https://www.fifatms.com/data-reports/reports/`

Frick, B. (2007). The football players'labor market: Empirical evidence from the major european leagues. *Scottish Journal of Political Economy*, *54*(3), 422–446.

Gerrard, B. (2001). A new approach to measuring player and team quality in professional team sports. *European Sport Management Quarterly*, *1*(3), 219–234.

Giuliani, M. (2013). Not all sunshine and roses: discovering intellectual liabilities "in action". *Journal of Intellectual Capital*, *14*(1), 127–144.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory*, *21*(1), 60–68.

He, M., Cachucho, R., & Knobbe, A. (2015). Football player's performance and market value. In *Proceedings of the 2nd workshop of sports analytics, european conference on machine learning and principles and practice of knowledge discov-ery in databases (ecml pkdd).*

Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, *17*(4), 484–492.

Jetter, M., & Ramírez Hassan, A. (2015). Want export diversification? educate the kids first. *Economic Inquiry*, *53*(4), 1765–1782.

Karl, A., & Lenkoski, A. (2012). Instrumental variable bayesian model averaging via conditional bayes factors. *arXiv preprint arXiv:1202.5846*.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Keidell, R. W. (1987). Team sports models as a generic organizational framework. *Human Relations*, *40*(9), 591–612.

Koop, G. (2003). *Bayesian econometrics.* Jhon Wiley and Sons Ltd,England.

Koop, G., León-Gonzalez, R., & Strachan, R. (2012). Bayesian model averaging in the instrumental variable regression model. *Journal of Econometrics*, *171*, 237–250.

Kuethe, T. H., & Motamed, M. (2010). Returns to stardom: evidence from us major league soccer. *Journal of Sports Economics*, *11*(5), 567–579.

Leamer, E. (1978). *Specification searches.* New York: New York: Weley.

Lenkoski, A., Eicher, T. S., & Raftery, A. (2014). Two-stage bayesian model averaging in endogeneous variable models. *Econometric Reviews*, *33*.

Ley, E., & Steel, M. F. (2009). On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of applied econometrics*, *24*(4), 651–674.

León-González, R., & Montolio, D. (2015). Endogeneity and panel data in growth regressions: A bayesian model averaging approach. *Journal of Macroeconomics*, *46*, 23–39.

Lucifora, C., & Simmons, R. (2003). Superstar effects in sport: Evidence from italian soccer. *Journal of Sports Economics*, *4*(1), 35–55.

Madigan, D., & York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 215–232.

Madigan, D., York, J., & Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 215–232.

Mollick, E., & Nanda, R. (2015). Wisdom or madness? comparing crowds with expert evaluation in funding the arts. *Management Science*, *62*(6), 1533–1553.

Montanari, F., Silvestri, G., & Bof, F. (2008). Performance and individual characteristics as predictors of pay levels: The case of the italian 'serie a'. *European Sport Management Quarterly*, *8*(1), 27–44.

Moral-Benito, E. (2013a). Determinants of economic growth: A bayesian panel data approach. *The Review of Economics and Statistics*, *94*(2), 566-579.

Moral-Benito, E. (2013b). Model averaging in economics: an overview. *Journal of Economic Surveys.*

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, *263*(2), 611–624.

Peeters, T. (2018). Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, *34*(1), 17–29.

Pennings, J. M., Lee, K., & Van Witteloostuijn, A. (1998). Human capital, social capital, and firm dissolution. *Academy of management Journal*, *41*(4), 425–440.

Pfeffer, J. (1994). Competitive advantage through people. *California management review*, *36*(2), 9–28.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual].

Vienna, Austria. Retrieved from `http://www.R-project.org/`

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997a). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997b). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191.

Raftery, A. E., & Zheng, Y. (2003). Discussion: Performance of bayesian model averaging. *Journal of the American Statistical Association*, *98*(464), 931–938.

Ramírez Hassan, A., & Graciano Londoño, M. (2017). Profiting from the English Premier League: Predictive elicitation, the Kelly criterion, and black swans. *International Journal of Sports Finance*, *12*, 386–395.

Sala-i Martin, X., Doppelhofer, G., & Miller, R. I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review*, 813–835.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, *26*, 393–415.

Scelles, N., Helleu, B., Durand, C., & Bonnal, L. (2016). Professional sports firm values: bringing new determinants to the foreground? a study of european soccer, 2005-2013. *Journal of Sports Economics*, *17*(7), 688–715.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Szymanski, S. (1997). Discrimination in the english soccer leagues: A market test. *Imperial College Management School Discussion Paper*.

Szymanski, S., & Smith, R. (1997). The english football industry: profit, performance and industrial structure. *International Review of Applied Economics*, *11*(1), 135–153.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

Weimar, D., & Wicker, P. (2017). Moneyball revisited: Effort and team performance in professional soccer. *Journal of Sports Economics*, *18*(2), 140–161.

Wright, P. M., Smart, D. L., & McMahan, G. C. (1995). Matches between human resources and strategy among ncaa basketball teams. *Academy of Management journal*, *38*(4), 1052–1074.

Zeugner, S. (2011). Bayesian model averaging with bms. *Tutorial to the R-package BMS 1e30*.

Appendix A

Technical appendix

Taking into account the normality assumption in stochastic errors, we have $p(\boldsymbol{y}|\boldsymbol{\theta}_r, M_r) \sim \mathcal{N}(\boldsymbol{i}_N\beta_0 + \boldsymbol{X}_r\boldsymbol{\beta}, h^{-1}\boldsymbol{I}_N)$. Given this likelihood and our prior distributions, we can use equation 3 to obtain the posterior distributions. In particular, the unconditional posterior distribution for $\boldsymbol{\beta}_r$ is multivariate student's t, that is, $\boldsymbol{\beta}_r|\boldsymbol{y} \sim \mathcal{T}(\bar{\boldsymbol{\beta}}_r, \bar{v}\bar{s}_r^2/(\bar{v}-2)\bar{\boldsymbol{V}}_r, N)$ where $\bar{\boldsymbol{\beta}}_r = \bar{\boldsymbol{V}}_r\boldsymbol{X}_r'\boldsymbol{y}$, $\bar{v} = N$, $\bar{s}_r^2 = \frac{\frac{1}{g_{k_r}+1}\boldsymbol{y}'\boldsymbol{P}_{\boldsymbol{X}_r}\boldsymbol{y} + \frac{g_{k_r}}{g_{k_r}+1}(\boldsymbol{y}-\bar{y}i_N)'(\boldsymbol{y}-\bar{y}i_N)}{\bar{v}}$, $\boldsymbol{P}_{\boldsymbol{X}_r} = \boldsymbol{I}_N - \boldsymbol{X}_r(\boldsymbol{X}_r'\boldsymbol{X}_r)^{-1}\boldsymbol{X}_r'$, and $\bar{\boldsymbol{V}}_r = [(1+g_{k_r})\boldsymbol{X}_r'\boldsymbol{X}_r]^{-1}$ (Koop, 2003).

An additional component that is crucial in *BMA* is the marginal likelihood,

$$
\begin{aligned}
p(\boldsymbol{y}|M_r) = & \int_{\Theta_r} p(\boldsymbol{y}|\boldsymbol{\theta}_r, M_r)p(\boldsymbol{\theta}_r|M_r)d\boldsymbol{\theta}_r \\
& \int_{\mathcal{R}^{++}}\int_{\mathcal{R}}\int_{\mathcal{R}^r} p(\boldsymbol{y}|\boldsymbol{\beta}_r, \beta_0, h, M_r)p(\boldsymbol{\beta}_r|h, M_r)p(\beta_0)p(h)\ d\boldsymbol{\beta}_r\ d\beta_0\ dh \\
& \propto \left(\frac{g_{k_r}}{g_{k_r}+1}\right)^{k_r/2}\left[\frac{1}{g_{k_r}+1}\boldsymbol{y}'P_{\boldsymbol{X}_r}\boldsymbol{y} + \frac{g_{k_r}}{g_{k_r}+1}(\boldsymbol{y}-\bar{y}\boldsymbol{i}_N)'(\boldsymbol{y}-\bar{y}\boldsymbol{i}_N)\right]^{-\frac{N-1}{2}}.
\end{aligned}
$$

We implement *BMA* using a birth/death *MCMC* algorithm, that is an adaptation of the mechanism developed by Madigan et al. (1995). In particular, given the space of models $\mathcal{M}$, we simulate a chain of M_s models, $s = 1, 2, ..., S << 2^k$, where the algorithm randomly extracts a candidate model M_c from a neighborhood of models ($nbd(M)$) that consists of the actual model itself, and the set of models with either one variable more or one variable fewer (Raftery et al., 1997b). Therefore, there is a transition kernel in the space of models $q(M \to M_c)$, such that $q(M \to M_c) = 0\ \forall M_c \notin nbd(M)$ and $q(M \to M_c) = \frac{1}{|nbd(M)|}\ \forall M \in nbd(M)$, $|nbd(M)|$ is the number of neighbors of M. This candidate model is accepted with probability

$$
\alpha(M_{s-1}, M_c) = Min\left\{\frac{|nbd(M)|p(\boldsymbol{y}|M_c)\pi(M_c)}{|nbd(M^c)|p(\boldsymbol{y}|M_{(s-1)})\pi(M_{(s-1)})}, 1\right\}.
$$

We address endogeneity issues in our approach using *IVBMA*, developed by Karl and Lenkoski (2012). We specify the dependent variable as a linear function of one endogenous regressor and exogenous regressors, that is $y_i = \boldsymbol{x}_{ei}'\boldsymbol{\beta}_1 + \beta_s x_{si} + \mu_i$ where $x_{si} = \boldsymbol{x}_{ei}'\gamma_1 + \boldsymbol{z}_i'\boldsymbol{\gamma}_2 + v_i$, $\boldsymbol{x}_s$ is the variable which generates endogeneity issues, such that $\boldsymbol{x}_e$ are k_1 exogenous regressors and $\boldsymbol{z}$ are k_2 instruments. Assuming $(u_i, v_i)' \overset{i.i.d}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = [\sigma_{lm}]$, $l, m = 1, 2$, the likelihood function is $\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\Sigma}|\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma})\boldsymbol{\Sigma}^{-1}\begin{pmatrix} y_i - \boldsymbol{x}_i'\boldsymbol{\beta} \\ x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma}\end{pmatrix}\right\}$ where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \beta_s]'$, $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2']'$, $\boldsymbol{x}_i = [\boldsymbol{x}_{ei}', x_{si}]'$ and $\boldsymbol{w}_i = [\boldsymbol{x}_{ei}', \boldsymbol{z}_i']'$.

Assuming $\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(3, \boldsymbol{I})$, the posterior distributions are

$$\beta|\boldsymbol{\gamma},\boldsymbol{\Sigma},\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{\beta}^*, \boldsymbol{B}^*)$$

$$\gamma|\boldsymbol{\beta},\boldsymbol{\Sigma},\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z} \sim \mathcal{N}(\gamma^*, \boldsymbol{G}^*)$$

$$\boldsymbol{\Sigma}^{-1}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z} \sim \mathcal{W}(\alpha^*, \boldsymbol{\Sigma}^*)$$

where $\boldsymbol{B}^* = (\omega_1^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' + \boldsymbol{I})^{-1}$, $\boldsymbol{\beta}^* = \boldsymbol{B}^*\left(\omega_1^{-1}\sum_{i=1}^{n}\left[\boldsymbol{x}_i\left(y_i - \frac{\sigma_{12}(x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma})}{\sigma_{22}}\right)\right]\right)$, $\omega_1 = \sigma_{11} - \sigma_{12}^2/\sigma_{22}$, $\boldsymbol{G}^* = (\omega_2^{-1}\sum_{i=1}^{n}\boldsymbol{w}_i\boldsymbol{w}_i' + \boldsymbol{I})^{-1}$, $\boldsymbol{\gamma}^* = \boldsymbol{G}^*\left(\omega_2^{-1}\sum_{i=1}^{n}\left[\boldsymbol{w}_i\left(x_{si} - \frac{\sigma_{12}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})}{\sigma_{11}}\right)\right]\right)$, $\omega_2 = \sigma_{22} - \sigma_{12}^2/\sigma_{11}$, $\alpha^* = 3 + n$ and $\boldsymbol{\Sigma}^* = \left[\boldsymbol{I} + \sum_{i=1}^{n}\begin{pmatrix} y_i - \boldsymbol{x}_i'\boldsymbol{\beta} \\ x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma}\end{pmatrix}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}, x_{si} - \boldsymbol{w}_i'\boldsymbol{\gamma})\right]^{-1}$.

Karl and Lenkoski (2012) propose an algorithm based on conditional Bayes factors that allows embedding MC^3 within a Gibbs sampling algorithm. Given the candidate (M_c^{2nd}) and actual (M_{s-1}^{2nd}) models for the iteration s in the second stage, the conditional Bayes factor is $CBF^{2nd} = \frac{p(\boldsymbol{y}|M_c^{2nd},\boldsymbol{\gamma},\boldsymbol{\Sigma})}{p(\boldsymbol{y}|M_{s-1}^{2nd},\boldsymbol{\gamma},\boldsymbol{\Sigma})}$, where $p(\boldsymbol{y}|M_c^{2nd},\boldsymbol{\gamma},\boldsymbol{\Sigma}) = \int_{\mathcal{M}^{2nd}} p(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\Sigma})\pi(\boldsymbol{\beta}|M_c^{2nd})d\boldsymbol{\beta} \propto |\boldsymbol{B}^*|^{1/2}Exp\left\{\frac{1}{2}\boldsymbol{\beta}^{*\prime}\boldsymbol{B}^{*-1}\boldsymbol{\beta}^*\right\}$. For the first stage, $CBF^{1st} = \frac{p(\boldsymbol{y}|M_c^{1st},\boldsymbol{\beta},\boldsymbol{\Sigma})}{p(\boldsymbol{y}|M_{s-1}^{1st},\boldsymbol{\beta},\boldsymbol{\Sigma})}$, where $p(\boldsymbol{y}|M_c^{1st},\boldsymbol{\beta},\boldsymbol{\Sigma}) = \int_{\mathcal{M}^{1st}} p(\boldsymbol{y}|\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\Sigma})\pi(\boldsymbol{\gamma}|M_c^{1st})d\boldsymbol{\gamma} \propto |\boldsymbol{G}^*|^{1/2}Exp\left\{\frac{1}{2}\boldsymbol{\gamma}^{*\prime}\boldsymbol{G}^{*-1}\boldsymbol{\gamma}^*\right\}$.[10] These conditional Bayes factors assume $\pi(M^{1st}, M^{2sd}) \propto 1$.

[10] In the case that $\beta_s = 0$, the update is based on seemingly unrelated regressions framework.

Appendix B

Variables and operationalization

Table B1

Variables

Variable name	Operationalization
Market value	Natural logarithm for the soccer player market value for the beginning of the season 2016-2017
Performance	Player appearances last season 2015-2016
Age	Age of athlete in years
National team	Dummy variable that takes the value of 1 if it has participated and 0 otherwise
Team position	Dummy variable for each position (goalkeeper, defender, midfield, and forward), taking the goalkeeper position as control level
Goals	Goals scored during career
Experience	Years since the athlete has been a soccer player for different teams (Barros, 2001)
Assist	Assist made by the player in the season 2015-2016
Nationality	City of birth according to classification of continental football associations (CAF, CONCACAF, CONMEBOL, OFC, AFC, and UEFA) (Kuethe & Motamed, 2010), taking the CAF association as control level
Goals previous season	Goals scored during 2015-2016 season (Gerrard, 2001)
League	Dummy for each leagues (Premier League, Bundesliga, LaLiga, Serie A and Ligue 1), taking LaLiga as control level
Coach	Number of games directed in the last season 2015-2016, in the league (Herm et al., 2014)
Matches	Matches played during the career in the league
Legs	Dummy variable for handling left leg, handling right leg, and both legs handling or ambidextrous (Bryson et al., 2013), taking the handling left leg as control level
Assistances matches	Assist as local (Lucifora & Simmons, 2003) in 2015-2016 season
Team performance	Dummy variable that takes the value of 1 if the athlete has a part of a prestigious soccer club in the season 2015-2016 and 0 otherwise (first division and second division). A club is prestigious if it had ranked at least fourth in the final ranking of its domestic league (Montanari et al., 2008)
Club administration	Sum of the market values of all team soccer players (Herm et al., 2014) in the 2015-2016 season, except the player that is analyzed
Sports agent	Dummy variable takes the value of 1 if the soccer player has agent and 0 otherwise
Height	Soccer player height in centimeters
Substitute	Appearances as a substitute (Bryson et al., 2013) in the 2015-2016 season
Public attention	Natural logarithm for the number of citations of the athlete in Google searcher (Herm et al., 2014)
Member union association	Dummy variable that takes the value of 1 if the league for which the player plays is part of the International Federation of Professional Football Players (FIFPro), and 0 otherwise
International players team	Percentage of foreign players in the team in the 2015-2016 season
Superstar 1	Dummy variable that takes the value of 1 if the player scores from 0.25 to 0.40 goals per match (Lucifora & Simmons, 2003) in the 2015-2016 season
Superstar 2	Dummy variable takes the value of 1 if the player scores more than 0.40 goals per match (Lucifora & Simmons, 2003) in 2015-2016 season
Player all-star game	Proxy. Dummy variable that takes the value of 1 if the player was chosen for the FIFA FIFPro World11 2016 award
Designated player	Proxy. Dummy variable takes the value 1 for the three most valuable soccer players for each team
Under-21 but not senior	Dummy variable takes the value of 1 if the player has been selected for national under 21 team but not for national senior team (Lucifora & Simmons, 2003) in the 2015-2016 season
Appearances UEFA champions	Player appearances in the UEFA champions League (Bryson et al., 2013) in the 2015-2016 season
Recognition under-21	Dummy variable that takes the value of 1 if it has participated in the national under 21 team and 0 otherwise
International recognition	Champion in the last tournament of the Confederations FIFA Cup, Euro UEFA of teams, Club World Cup, UEFA Champions League, UEFA European League, and UEFA Super Cup
Sportsman reputation	Number of seasons that the athlete has played for a prestigious club in the last five years or seasons (since 2011-2012 season to 2015-2016 season). Variable rated from 0 to 5, for that it is a prestigious club which has been in at least the top four positions in the local league (Montanari et al., 2008)
Team experience	Number of seasons the team has played in the first division league in the last five years (since 2011-2012 season to 2015-2016 season). Variable coded from 0 to 5 (Montanari et al., 2008)
Change of team	Dummy variable that takes the value of 1 if the athlete changes from club, during the 2015-2016 season, and 0 otherwise (Montanari et al., 2008)

Appendix C

Descriptive statistics

Table C1

Descriptive statistics

Variable	Mean	Std.Dev	Min.	Max.	Mode	Median
Market value (millions)	6.614	10.22	0.05	80	2	2.5
Performance	14.58	13.75	0	38	0	13
Age	26.14	4.322	17	38	28	26
National team	0.546	0.499	0	1	1	1
Goalkeeper	0.110	0.476	0	1	0	0
Defender	0.346	0.477	0	1	0	0
Midfield	0.269	0.444	0	1	0	0
Forward	0.275	0.447	0	1	0	0
Goals	26.22	36.48	0	279	0	13
Experience	9.436	4.446	0.30	23.02	10.01	9.01
Assists	1.493	3.533	0	30	0	0
Caf	0.113	0.318	0	1	0	0
Concacaf	0.015	0.121	0	1	0	0
Conmebol	0.182	0.387	0	1	0	0
Afc	0.015	0.121	0	1	0	0
UEFA	0.675	0.469	0	1	1	1
Goals previous season	1.072	2.039	0	15	0	0
Premier league	0.200	0.401	0	1	0	0
Bundesliga	0.194	0.396	0	1	0	0
LaLiga	0.188	0.391	0	1	0	0
Serie A	0.203	0.403	0	1	0	0
Ligue 1	0.215	0.411	0	1	0	0
Matches	86.92	89.81	0	446	0	60
Left leg	0.340	0.475	0	1	0	0
Right leg	0.636	0.482	0	1	1	1
Ambidextrous	0.024	0.153	0	1	0	0
Assistances matches (thousands)	547.7	312.6	98.8	1489	1139	459.3
First division	0.203	0.403	0	1	0	0
Second division	0.266	0.803	0	4	0	0
Club administration	141.6	140.3	9.85	701	408.6	80.8
Sport agent	0.797	0.403	0	1	1	1
Height	1.827	0.064	1.63	1.97	1.81	1.83
Substitute	3.573	5.923	0	38	0	1
Public attention (millions)	3.845	12.98	0.02	125	0.42	0.39
Member union association	0.606	0.489	0	1	1	1
International players team	0.550	0.175	0.04	1	0.75	0.56
Superstar 2	0.003	0.055	0	1	0	0
Player all-start game	0.003	0.055	0	1	0	0
Under-21 but not senior	0.015	0.121	0	1	0	0
Appearances UEFA Champions	0.143	0.351	0	1	0	0
Recognition under-21	0.463	0.499	0	1	0	0
Fifa cup	0.003	0.055	0	1	0	0
Euro UEFA	0.003	0.055	0	1	0	0
UEFA champion	0.003	0.055	0	1	0	0
UEFA European	0.009	0.094	0	1	0	0
UEFA super cup	0.003	0.055	0	1	0	0
Team experience	2.418	1.948	0	5	5	2
Coach	25.56	14.10	0	38	38	34
Change of team	0.218	0.413	0	1	0	0