Strategy and Business Economics

Ticket consumption forecast for Brazilian championship games

Previsão de consumo de ingressos para jogos do campeonato brasileiro

Estimación de consumo de entradas de partidos del campeonato brasileño

Adriana Bruscato Bortoluzzo [a,*], Mauricio Mesquita Bortoluzzo [b], Sérgio Jurandyr Machado [c], Tatiana Terabayashi Melhado [a], Pedro Iaropoli Trindade [a], Bruno Santos Pereira [a]

[a] *Insper Instituto de Ensino e Pesquisa, São Paulo, SP, Brazil*
[b] *Universidade Presbiteriana Mackenzie, São Paulo, SP, Brazil*
[c] *Secretaria de Política Econômica, Brasília, DF, Brazil*

Abstract

For the efficiency of sales and marketing management of athletic clubs, it is crucial to find a way to appropriately estimate the level of demand for sporting events. More precise estimates allow for an appropriate financial and operational plan and a higher quality of service delivered to the fans. The focus of this study is to analyze and forecast the ticket consumption for soccer games in Brazilian stadiums. We compare the results of the regression model with normally distributed errors (benchmark), the TOBIT model and the Gamma generalized linear model. The models include explanatory variables related to the economic environment, product quality, as well as monetary and non-monetary incentives that people are given to attend sporting events at stadiums. We show that most of these variables are statistically significant to explain the amount of fans that go to stadiums. We used different measures of accuracy to evaluate the performance of demand forecasts and concluded that Gamma generalized linear model presented better results to forecast the ticket consumption for Brazilian championship games, when compared to a benchmark.
© 2016 Departamento de Administração, Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo – FEA/USP. Published by Elsevier Editora Ltda. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Sports management; GLM; Ticket consumption; Soccer

Resumo

A estimação da demanda em eventos esportivos é uma questão crucial para a avaliação da eficiência de vendas e gestão de marketing de clubes desportivos. Estimativas mais precisas permitem que seja feito um plano financeiro e operacional mais adequado e que o serviço prestado aos fãs possua maior qualidade. O foco deste estudo é analisar e prever o consumo de ingressos em jogos de futebol nos estádios brasileiros. Foram comparados os resultados preditivos do modelo de regressão com erros normalmente distribuídos (*benchmark*), do modelo Tobit e do modelo linear generalizado com distribuição Gama. Os modelos incluem variáveis explicativas relacionadas com o ambiente econômico, a qualidade do produto, bem como incentivos monetários e não monetários que as pessoas possuem para assistirem aos jogos nos estádios. A maioria dessas variáveis foi estatisticamente relevante para explicar a quantidade de público pagante nos estádios. Foram utilizadas diferentes medidas de qualidade de previsão

---

* Corresponding author at: Rua Quatá 300, 04546042 São Paulo, SP, Brazil.
  E-mail: adrianab@insper.edu.br (A.B. Bortoluzzo).

para avaliar o desempenho das previsões de demanda e concluímos que o modelo linear generalizado com distribuição Gama apresentou melhores resultados para previsão do consumo de ingressos para os jogos do Campeonato Brasileiro de futebol, quando comparados ao *benchmark*.

Resumen

La estimación de la demanda en eventos deportivos es un tema crucial para la evaluación de la eficiencia de ventas y gestión de marketing de clubes deportivos. Estimaciones más precisas permiten que se cree un plan financiero y operativo más adecuado y que se ofrezca un servicio de mejor calidad a los aficionados. El objetivo de este estudio es analizar y estimar el consumo de entradas para partidos de fútbol en los estadios brasileños. Se han comparado los resultados del modelo de regresión con errores normalmente distribuidos (*benchmark*), del modelo Tobit y del modelo lineal generalizado con distribución Gamma. Los modelos incluyen variables explicativas relacionadas con el entorno económico, la calidad del producto y los incentivos monetarios y no monetarios que se ofrecen a las personas para que asistan a los partidos en los estadios. Se demuestra que la mayoría de estas variables ha sido estadísticamente significativa para explicar la cantidad de personas que pagan para ir a los estadios. Se han utilizado diferentes medidas de calidad de estimación para evaluar el desempeño de las previsiones de demanda y se ha concluido que el modelo lineal generalizado con distribución Gamma muestra mejores resultados para estimar el consumo de entradas de los partidos de fútbol del Campeonato Brasileño, en comparación con el *benchmark*.

Introduction

Soccer plays an undeniably important role not only in the context of Brazilian sports, but also in the world of international sports. Nevertheless, Brazilian soccer clubs, some of them at over 100 years old, have many organizational problems and are, in general, overwhelmed with chronic mismanagement. The professionalization of the sport and its command structures are weak and lag behind the levels of organization and development achieved by its European counterparts.

As the most popular sport in Brazil, 5.7 million fans attended soccer games in Brazilian stadiums during the 2013 season – according to the Brazilian Soccer Confederation (CBF) – an amount that could be considered meager when compared to the 13.6 million fans that attended games in Premier League stadiums during the same season (2012–2013). The flood of European fans to their stadiums, especially since the 1990s, is a direct result of the high organizational standards of the European leagues (Sloane, 1997). These high standards maintained in Europe allow the major clubs to reach maximum attendance capacity in their stadiums for virtually every game of the season. In Brazil, by contrast, the number of fans attending games has been declining over the past decades (Giovannetti, Rocha, Sanches, & Silva, 2006).

Planning is always based on certain assumptions about the future course of events. Future conditions are often difficult to forecast, and can never be predicted perfectly. Yet, the marketer or the administrator must plan and make decisions using what constitutes the best estimate about future developments. Without a proper consumption forecast, the marketing executive cannot determine the type of marketing program to use in order to attain the desired sales and marketing objectives (Santos, Bazanini, & Ferreira, 2014). Therefore, evaluating the consumption potential and preparing a consumption forecast is an important function of sales and marketing managers. Mentzer and Moon (2005) define demand forecast as "a projection into the future of expected demand, given a stated set of environmental conditions". According to them, one of the key measures of sales forecasting performance is the accuracy of the forecast. For this purpose, it is essential to identify and understand the factors, both positive and negative, that influence attendance at sporting events. The aforementioned analyses will prove instrumental in Brazilian clubs as these clubs seek to increase attendance, and will thus contribute to more efficient and professional management of the sport. The focus of this study, therefore, is to examine the ticket consumption for soccer stadiums in Brazil through an analysis of the paying public of the Campeonato Brasileiro Série A (the A league division of the Brazilian championship) between 2004 and 2013. This study is one of the longest investigations ever conducted in sports literature with soccer data.

First, it is necessary to determine which variables influence the consumption of games by using regression models. Three models were fitted: a usual OLS regression model with normally distributed errors; a regression model with a censored dependent variable (TOBIT); the number of tickets sold, which is limited by the capacity of the stadium where each game is played, as proposed by Falter, Pérignon, and Vercruysse (2008); and generalized linear models (GLM) with Gamma distribution to better adjust the positive skewness of the consumption distribution. We will check if the TOBIT and the GLM models could be better to forecast ticket consumption than the usual regression models that have been used in the Brazilian literature (Giovannetti et al., 2006; Madalozzo & Villar, 2009; Melo, 2007; Souza, 2004), which does not use this specific tooling to estimate the demand for soccer in Brazil. Our intention is to fill this gap by providing more accurate estimates for the decision-making process within the ambit of Brazilian soccer clubs, allowing for better in-sample and out-sample demand forecasts.

The objective of this study is to contribute to the sports management literature in two ways. First, our framework integrates different types of explanatory variables to explain the ticket consumption of soccer games in Brazil, which include those related to the economic environment, product quality and monetary and non-monetary incentives that people are given to attend Sporting events at stadiums. For managers, our paper elucidates the relevant measures that will become the focal point of a club's effort to leverage its game revenues. Second, we produce more reliable forecasts of soccer ticket consumption by using generalized regression models that considers both the positive skewness of consumption distribution, and the restriction due to the capacity of stadiums.

The remainder of this paper is divided into the following sections: Section "Literature review and the description of study variables" presents the literature review and description of variables incorporated in the model based on the theoretical framework, Section "Statistical methodology" briefly describes the methodology, Section "Results" presents the interpretation of results, and Section "Conclusion" presents conclusions and suggestions for future research.

Literature review and the description of study variables

A merely superficial analysis of the Brazilian soccer management process is sufficient to detect the indelible need for professional and organizational development aimed to increase the efficiency of management (Melo, 2007). According to Park, Lee, and Miller (2012), sports teams have three main sources of revenue: ticket sales, sponsorship and the sale of broadcast rights. A number of factors can influence the demand for sports, including ticket prices, fans' income and wealth, population density near the stadium, the quality of the teams and the infrastructure of the stadiums where the matches occur. Thus, it is important for clubs to study how these variables affect the demand for championship games, so they are able to make reliably predictions of soccer ticket revenues.

One of the crucial issues of effective management is embodied in the estimation of the levels of consumption for those sporting events in which a particular club will participate throughout the year. The more precise the estimate, the better equipped management is to appropriately plan the organization's financial and operational needs. A better plan may result in a higher quality of service delivery to the fans, and, why not, in a better performance of the club on the field. Together, these two factors lead to a greater number of fans in the stadium and, consequently, a revenue increase generated by the event, thus forming a profitable cycle.

As pointed out by Smith and Groetzinger (2010, p. 4), "it is possible it would be profitable for teams to drop prices for the purpose of increasing the chance of victory, as the revenue associated with an additional win could outweigh lost ticket revenue". Even if that is the case, it is essential to understand consumers' demand for soccer games (Alonso & O'Shea, 2013). García and Rodríguez (2002) specified a consumption equation with the use of economic variables and proxies to control factors such as quality of the match, unpredictability of the

outcome of the game, and the opportunity cost for the match. Falter and Pérignon (2000) divided the explanatory variables of the consumption for sports events into three groups: variables related to (i) the economic environment, (ii) product quality, and (iii) incentives to go to the stadium. We understand that at this stage of the literature, the grouping of the explanatory variables between factors is still an arbitrary question, especially because these two last papers had worked with similar variables in the end. Here we opted to use the same grouping done in Falter and Pérignon (2000) to an easy understand of the problem.

The group of economic environment variables, also called structural variables, affects the consumption of the good (the purchase of tickets), which affects the consumer's budget and is limited by available income. Therefore, two proxies were chosen to represent the economic environment: per capita income in the city where the game occurs (PCI) and the population of the city (POP). The latter was included because a greater influx of people is expected in those stadiums located in the most densely populated cities. Data on these variables were obtained from the website of the Brazilian Institute of Geography and Statistics [IBGE] (http://www.ibge.gov.br, retrieved in 25, February, 2015).

Income elasticity of consumption measures the sensitivity of the consumption of a good related to changes in consumer income, ceteris paribus A negative income elasticity of ticket consumption is associated with inferior goods while a positive one is associated with normal goods. If income elasticity of a normal good is less than unity, it is a necessity good, and if income elasticity is greater than unity, it is a luxury or a superior good. Bird (1982) found that an increase in household income causes a drop in demand for tickets in English soccer, suggesting that soccer match tickets are an inferior good. Madalozzo and Villar (2009) found a negative relationship between the average per capita income and ticket consumption in Brazilian games.

The second group includes those variables related to product quality, that is, the quality of the soccer team that sells its tickets and the quality of the opponent. These data seek to measure the performance of the home and visiting teams in the tournament.

Szymanski (2001) examined 997 games over 22 years between the same teams in both the English Premiership and FA Cup. The author states that the sum of the position of participating teams (in the championship) reflects statistically significant influence on the consumption of tickets; that is, better positioned teams in the championship are more likely to have an increased number of fans in the stadiums.

Thus, the importance of the match for both home and visiting teams is relevant to understanding ticket consumption. This factor will be represented in the study by the teams' positions in the league when those teams are facing each other in a match. Variables that will be important for the reproduction of the match will be the position of the home team (CLH) and the position of the visiting team (CLV) in the league on the date of the game.

In addition to position in the tournament, the model includes variables associated with goal differences and results in recent matches. The variables that express the number of points earned in the last three games before the match in question by the home

team (PGH) and by the visiting team (PGV) were added to the model to measure the performances of the team right before the match. This stems from the premise that teams with higher values for these variables demonstrate improved quality and efficiency.

Two other variables that seek to measure the expected quality of games are added to the model. These variables represent the sum of the goals scored by the home team (GLH) and the visiting team (GLV), in the three previous rounds to the game in question. It is implied that the greater the number of goals scored, the higher the quality of the match, as goal scoring is the part of the game the fans enjoy and appreciate most. Games with more goals tend to be viewed as more entertaining games, as pointed out by Calster, Smits, and Huffel (2008).

The importance of the match in the league can also affect ticket consumption. As a championship series progresses, the higher the relative value of the game and, therefore, the more attendees. For this reason, a given target league will be divided into two phases for analysis; furthermore, the beginning and the ending of each phase will also be considered. Thus, the tournament will be divided into four parts to differentiate the degree of relative importance of matches that occur at different times during the season. We will have variables indicating parts 2 (PT2), 3 (PT3) and 4 (PT4), and it is expected that as the championship rounds advance, public attendance will increase.

The rivalry between teams is directly related to the degree of importance of a match. The higher the equilibrium and rivalry between the teams, the greater the interest of the fans in the match; thus, there should be an increase in the ticket consumption for the game. Aiming to represent the effect of rivalry between teams, a dummy variable was included between the most traditional teams of the same state (CLS). Due to historically high competitiveness between two in-state teams and the great rivalry between the fans, these matches are expected to be of high quality (Wooten, March 5, 2015).

A variable to represent the presence of a big team from Sao Paulo or Rio de Janeiro as a guest of a match (BIG) has also been added to the model because these teams have great historical importance in Brazilian soccer and have many fans throughout the country. This variable assumes a value 1 when a home team that is not from Sao Paulo or Rio de Janeiro faces such teams as Corinthians, Palmeiras, Sao Paulo, Santos, Flamengo, Fluminense, Botafogo or Vasco, and assumes a value of zero for other cases. It is expected that the presence of big clubs from these two states will increase ticket consumption.

The third group of variables represents the incentives that fans are given to attend a soccer match at the stadium. There is a significant monetary incentive associated with ticket prices. Accordingly, a variable representing the average ticket price (PRC) was included, and it was obtained by dividing the income earned in the game by the number of paying attendees. When price elasticity of the consumption of a good is elastic, lowering its price causes an increase in revenues, and when its consumption is inelastic revenues fall. Several studies found evidence of inelastic consumption for tickets across sports and countries: Jennett (1984) in the Scottish soccer league, Borland and Lye (1992) in the Australian Rules soccer, García and Rodríguez

(2002) in the Spanish soccer league, Madalozzo and Villar (2009) for Brazilian soccer league, and Nilon (2010) for English Premier League.

According to Kotler (1975), nonprofit organizations do not seek to stablish a price to profit maximization; instead, they try to stablish a "fair" price, according to its costs. Furthermore, there are practical problems in using the profit-maximizing price among which we highlight: the trade-off between short-term and long-term profit; possibility of boycott of supporters and; impossibility of accurately estimate the demand.

In Brazil, football firms (hooligans), and directors attached to them, tend to put pressure on the manager in the establishment of a less austere pricing policy. A high price could then cause various problems to the manager. In the short term, he would have to deal with public dissatisfaction and possible loss of mandate, while in the long run he can have a great reduction of supporters (market share). Therefore, we believe that most of the time the leaders are more concerned with charging a price that covers the operating costs of the stadium than in earning income from matches.

At least in the short term, the football clubs are monopolists in home matches. They are the only suppliers and set the price of a ticket departure of his team. As one can consider, the marginal cost of a ticket sold is zero (all costs of a football stadium are fixed), the monopoly profit maximization problem comes down to maximize revenue in that match. This requires selling the ticket at a price where demand has a unitary elasticity. If demand is elastic, than the leader can increase the number of viewers and profit while only lowering the price of admission, which would lead to satisfaction of fans and increase profit at the same time. However, if demand is inelastic, the profit increase will only occur with an increase in price and dissatisfaction of the fans. So if the estimated demand is inelastic, we have strong evidence that the leaders are more concerned with the satisfaction of fans and less with profit.

There are also several "non-monetary incentives". According to Knowles, Sherony, and Haupert (1992) and Simmons (1996), the time and day of the week in which games occur have a significant influence on ticket consumption. The authors demonstrate that matches held during the evening are more attractive to the public than daytime games held during the week. Weekend games are even more attractive than those played in the evenings on weekdays. Variables used to convey these desired effects are games on weekends (WND) and games on weekdays after 9:00 PM (NGT).

The weather on game day may also have an important explanatory effect with respect to the number of fans who go to the stadium. Such factors as extremely high or low temperatures, or rainy days can be major factors that influence attendees' decision to attend a match. To imprint this effect into the study, we included a variable that represents the accumulated precipitation in mm on a game day and in the city where the match occurred (RAN). It can be expected that the more total rainfall recorded on a game day, the lower the incentive for fans to attend a game depending on (i) the poor infrastructure of the stadiums in Brazil (for example, not all seats are covered or protected

Table 1
List of explanatory variables and the expected effect.

Variable	Description	Expected effect
Economic enviroment variables		
PCI	Annual per capita income in the city where the game takes place (in R$)	–
POP	City population where the game occurs	+
Variables related to product quality (soccer specific variables)		
CLH	Classification of the home team	–
CLV	Classification of the visiting team	–
PGH	Points won by the home team in the past 3 games	+
PGV	Points won by the visiting team in the past 3 games	+
GLH	Goals scored by the home team in the past 3 games	+
GLV	Goals scored by the visiting team in the past 3 games	+
PT2	Value 1 if the match occurs on the 2nd stage of the championship	+
PT3	Value 1 if the match occurs on the 3rd stage of the championship	+
PT4	Value 1 if the match occurs on the 4th stage of the championship	+
CLS	Value 1 if the match is considered a classic	+
BIG	Value 1 if the game has a major league team from SP or RJ	+
Incentive variables		
PRC	Average ticket price	–
WND	Value 1 if the match occurred on the weekend	+
NGT	Value 1 if the match occurred after 9:00 PM	+
RAN	Rainfall (in mm)	–

from the rain in most Brazilian stadiums), (ii) the expected drop in quality of a match when there is rain, and (iii) the difficulties associated with transportation to the stadium on a rainy day.

Considering the literature review, we understand that there are 3 constructs that affect the paying public, presented in Fig. 1, and as the constructs are multidimensional, we choose more than one proxy in the literature in order to better represent each one of them. Thus we aim to test 3 hypotheses:

H1. The economic environment has significant influence over the paying public.

H2. A higher quality product implies a higher paying public.

H3. Greater incentives entail a higher paying public.

These hypotheses can be confirmed if we find the expected signs of the variable coefficients and statistical significance, such as summarized in Table 1. The table describes the variables that could explain the paying public and the expected effect of each of them on the consumption according to literature review.

Data concerning the average ticket price, day of the week, time and location of the matches were obtained through consultations and examinations of overviews and bordereau

on the website of the Brazilian Soccer Confederation, as well as from tables provided by the Placar magazine's website. Data for cumulative amounts of precipitation were obtained from the National Institute of Meteorology [INMET] (http://www.inmet.gov.br/portal/, retrieved in 15, may, 2015).

The unit of analysis is the soccer match and because the cross-sectional dataset spans the years 2004–2013, we will include variables indicating the year of the matches to control the possible effect of time in ticket consumption.

Finally, one explanatory variable related to the television broadcast of the soccer match must be factored into the equation, as Allan and Roy (2008) and Cox (2012) have warned that such transmissions negatively affect the number of fans who go to the stadium to watch the game. Baimbridge, Cameron, and Dawnson (1996) estimate an approximate 15% reduction in attendance of the *English Premier League* games that are televised during the week. In Brazil, as noted by Madalozzo and Villar (2009), "in the study period, the TV (open and by subscription) transmitted games in all rounds for all of Brazil, respecting the concept of not transmitting the match to the city where it was played. These games were available only on *pay-per-view*". In specific cases, as in the final games of the championship, there were situations where certain games were transmitted to the cities in which the games were held if the stadium sellout capacity was reached.

Statistical methodology

The statistical approach used so far by studies on the Brazilian context is not ideal, as it neglects the fact that there is a restriction due to the capacity of stadiums and that the consumption has a positive skewness. Aiming to account for important features of the distribution of ticket consumption and to make better forecasts, we decided to work with more general regression models than the usual multiple linear regression model with normal distribution for the errors. We use the TOBIT model that takes into account that the observed consumption is censored by the capacity of the stadium and the generalized linear regression model (GLM) with Gamma and Inverse Gaussian distributions for the response variable in order to adjust the positive skewness of ticket consumption.

The TOBIT regression model was proposed by Tobin (1958) and the advantage of using this model is that it allows us to work with censored variables; that is, where the real value is not noticed. In this work, the real ticket consumption has not been observed for games where tickets were sold out; that is, the consumption was censored and the observed value corresponds to the limit of tickets for sale (stadium sold out). If the regression is based only on the sample of games that are not at capacity, the ordinary least squares estimators (OLS) of parameters will be biased and also inconsistent (Wooldridge, 2002, p. 524). Consequently, marginal effects calculated based on the OLS estimators would not be suitable for the real situation and, ergo, would lead to incorrect interpretations of the impact of each of the variables on the consumption.
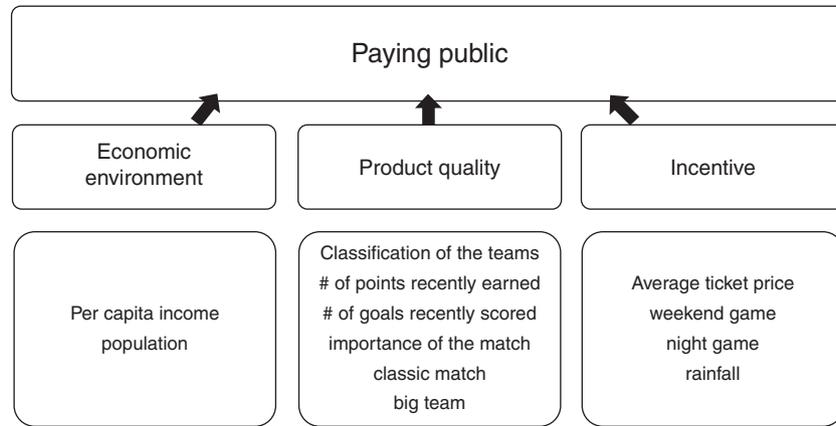
Fig. 1. Multidimensional constructs that affect the paying public and the selected proxies for them.

Table 2
Descriptive analysis of numerical variables.

Variable	n	Average	Standard deviation	Minimum	Maximum	Skewness
PUB	3659	13,939	11,033	147	87,895	1.89
LOT	3660	42,259	22,525	9500	95,000	0.78
PCI	3660	2262	584	364	3835	−0.17
POP	3639	3,608,558	3,771,287	15,051	11,253,503	1.10
CLH	3660	10.57	6.14	0	24	0.07
CLV	3660	10.55	6.13	0	24	0.07
PGH	3602	3.83	2.31	0	10	0.25
PGV	3602	4.16	2.34	0	12	0.15
GLH	3601	3.89	2.14	0	13	0.45
GLV	3597	4.10	2.20	0	14	0.44
PRC	3583	20.36	12.27	1.00	39,099	10.08
RAN	3660	3.43	9.47	0	110	4.37

Thus, in our case the TOBIT model expresses the observed response, Y, in terms of an underlying latent variable:

$$Y^* = \mathbf{X}\boldsymbol{\beta} + \varepsilon \quad \text{and} \quad Y = \min(\text{LOT}, Y^*),$$

where $\mathbf{X}$ is the information matrix, $\boldsymbol{\beta}$ is the vector of parameters, ε is the random error, min is the function minimum and LOT is the maximum capacity of the stadium. The model satisfies the same assumptions of the usual Normal regression model: linearity in parameters, random sampling, zero error conditional mean, error homoskedasticity, no perfect collinearity and errors with Normal distribution. The estimation is done using the method of maximum likelihood. However, due to the censure observed, the process involves a weighting of uncensored with censored variables in order to eliminate the bias of estimators and adapt the method for latent models. Then the log-likelihood for each observation is written as

$$l_i\left(\boldsymbol{\beta}, \sigma | Y_i, \mathbf{X}_i\right) = I\left(Y_i = \text{LOT}\right) \log\left(1 - \Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right)$$

$$+ \text{I}\left(Y_i < \text{LOT}\right) \log\left(\frac{1}{\sigma}\phi\left(Y_i - \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right), \qquad (1)$$

where σ is the standard deviation for ε, I is the indicator function, Φ and ϕ are the cumulative standard Normal distribution function and the standard Normal density function, respectively. The log-likelihood for a random sample of size n is obtained by

summing (1) across all i. The maximization of the log-likelihood requires numerical methods and we used the Eviews package to obtain the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ.

More details about the TOBIT, its estimation and asymptotic properties of estimators can be found in Wooldridge (2002), Greene (2003) and Gujarati (2004).

The idea of GLM is to expand the range of options for the error distribution used in the linear regression model, as well as provide greater flexibility to the functional relationship between the average of the response variable and the linear predictor. Therefore, the response variable can have any distribution that belongs to the exponential family, like Poisson, Gamma, Inverse Normal, Binomial, among others, leading to a better fit of the regression models, since they work with distributions that have different characteristics compared to the normal distribution.

Generalized linear models are characterized by the distribution of the response variable and the relation of the average of this variable and the linear predictors

$$E\left(Y_i\right) = \mu_i = g^{-1}\left(\mathbf{X}_i\boldsymbol{\beta}\right), \qquad (2)$$

where $\mathbf{X}_i$ is the linear predictor (or information matrix) and g is a monotonic and differentiable function, called link function.

There are several possibilities of distributions that could be used and we choose two different distributions: Gamma and Inverse Gaussian. The Gamma and the Inverse Gaussian
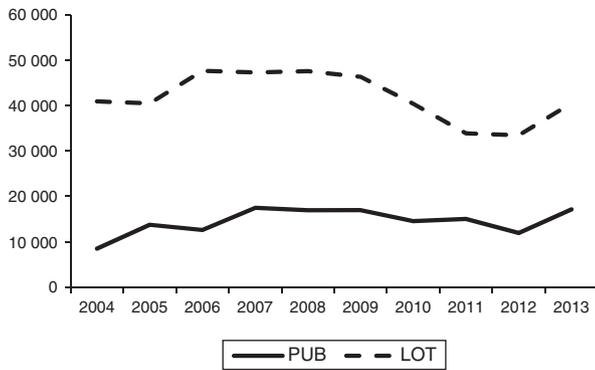
Fig. 2. Evolution of the annual average paying public (PUB) and the average capacity of stadiums (LOT) in the Brazilian championship from 2004 to 2013.

Table 3
Descriptive analysis of categorical variables.

Variable	Categories	Frequencies	Relative frequencies (%)
CLS	Yes	292	8.0
	No	3368	92.0
BIG	Yes	1683	46.0
	No	1976	54.0
WND	Yes	2609	71.3
	No	1051	28.7
NGT	Yes	413	11.3
	No	3247	88.7
PT	1	847	23.1
	2	971	26.5
	3	955	26.1
	4	887	24.2

distributions are both positive skewed and we consider the log link function, that is, the mean and variance depend on the explanatory variables through

$$\text{Gamma}: \quad E(Y_i) = \mu_i = \exp(\mathbf{X}_i\boldsymbol{\beta}) \text{ and } Var(Y_i) = \mu_i^2. \quad (3)$$

$$\text{Inverse Gaussian}: \quad E(Y_i) = \mu_i$$
$$= \exp(\mathbf{X}_i\boldsymbol{\beta}) \text{ and } Var(Y_i) = \mu_i^3. \quad (4)$$

We used the maximum likelihood method to estimate the parameters of GLM with the Quadratic Hill Climbing optimization algorithm. The log-likelihood for each observation is

$$l_i(\boldsymbol{\beta}, \sigma | Y_i, \mathbf{X}_i) = \frac{(Y_i\theta_i - b(\theta_i))}{\tau} + c(Y_i, \tau), \quad (5)$$

where τ is the dispersion parameter, typically is known and is usually related to the variance of the distribution, θ is the canonical parameter, in general equal to μ, and c is a constant that depends on the distribution of Y. More details about the GLM, its estimation and asymptotic properties of estimators can be found in McCullagh and Nelder (1989).

Results

As mentioned earlier, the dependent variable (PUB) is the paying public of the games analyzed. The cross-sectional dataset is obtained from summaries and bordereau on the website of the Brazilian Soccer Confederation (CBF) from 2004 to 2013 and the unit of analysis is the soccer match. In total, there were 3660 matches, but only 115 (3.1%) of which had maximum capacity and were regarded as censored in the TOBIT model.

Fig. 2 shows the graph of the evolution of the annual average paying public and the annual average capacity of the stadiums in the Brazilian championship from 2004 to 2013. The average occupancy rate was 34% of full capacity during this period, a low percentage, especially when compared with European standards, which in the 2013–2014 Premier League season had an average occupancy rate of 95%. It is noteworthy that there was an increase of approximately 18% in the capacity of stadiums from 2005 to 2006, whereas capacity remained constant from 2006 to 2009, and there was a decrease of 27% in the average

capacity of stadiums from 2009 to 2011. In terms of audience, there was a 62% increase from 2004 to 2005 and a 40% increase from 2006 to 2007; no major changes were indicated from 2007 to 2011. Due to this increase in audiences over the assessed time, we include year dummy variables in the model.

Table 2 presents the descriptive analysis of numerical variables used in the study and described in Table 1. Note that the average attendance in the stadiums during the study period is 13,939 fans per game with a high dispersion. The average maximum capacity of the stadiums in the Brazilian championship games during the specified period was 42,260, more than three times the paying public. There is great variability in these values with a minimum of 9500 and a maximum of 95,000. As already pointed out, the positive skewness of the paying public guided the choice of the generalized linear model (GLM) with Gamma and Inverse Gaussian distributions.

In the descriptive analysis of the categorical variables presented in Table 3, we can note that 8.0% of the games were considered a classic, 46.0% of the games include the major teams from Sao Paulo or Rio de Janeiro as visitors, 71.3% of the games occurred on weekends and 11.3% of the games took place after 9:00 PM. Also, we verify that there is no strong multicollinearity among explanatory variables as pointed by its correlation coefficients (Table 4).

The model estimation used 3260 observations, which 3175 were complete, and 400 observations were left to assess the out-of-sample forecasts. We then compared the accuracy of the forecasts of the in-sample estimated model using the in-sample and out-of-sample observations.

We use three specifications for the regression model that has the natural logarithm of paying public as the response variable: usual linear regression model (LRM) with normally distributed errors, (1) TOBIT model truncated by the maximum capacity of the stadiums (LOT), and generalized linear model (GLM) using (3) Gamma and (4) Inverse Gaussian distributions. The estimated model coefficients and their standard errors are in Table 5. For each model, an analysis of residuals was conducted to verify the assumptions about the errors and the necessary corrections have been implemented. In Table 5 we presented only the results of Gamma GLM because this model had less sum of squares than

Table 4
Pearson correlation for numerical variables.

	PUB	LOT	PCI	POP	CLH	CLV	PGH	PGV	GLH	GLV	PRC	RAN
PUB	1.000	0.382	−0.002	0.217	−0.241	−0.098	0.247	0.069	0.184	0.076	−0.347	−0.037
LOT		1.000	0.119	0.297	−0.111	−0.013	0.064	0.008	0.058	0.046	0.006	−0.022
PCI			1.000	0.086	−0.124	0.001	0.094	−0.016	0.151	0.073	−0.139	−0.019
POP				1.000	−0.176	0.007	0.080	−0.019	0.044	0.000	0.230	−0.045
CLH					1.000	−0.016	0.056	−0.348	−0.285	−0.019	−0.151	0.028
CLV						1.000	−0.034	−0.407	−0.029	−0.256	−0.043	0.003
PGH							1.000	0.148	0.541	0.028	0.079	−0.006
PGV								1.000	0.052	0.536	0.023	−0.013
GLH									1.000	0.120	0.028	0.002
GLV										1.000	−0.021	−0.014
PRC											1.000	−0.038
RAN												1.000

Table 5
Estimated regression models for ticket consumption in the Brazilian championship games.

Variable	LRM		TOBIT		Gamma-GLM	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
Log PCI	−0.207[**]	0.064	−0.262[**]	0.062	−0.237[**]	0.054
Log POP	0.166[**]	0.012	0.164[**]	0.011	0.151[**]	0.010
CLH	−0.019[**]	0.002	−0.020[**]	0.002	−0.017[**]	0.002
CLV	−0.012[**]	0.002	−0.013[**]	0.002	−0.009[**]	0.002
PGH	0.048[**]	0.006	0.049[**]	0.007	0.043[**]	0.006
PGV	−0.005	0.006	−0.004	0.007	−0.008	0.006
GLH	0.025[**]	0.007	0.027[**]	0.007	0.022[**]	0.006
GLV	0.014[*]	0.007	0.013[*]	0.007	0.017[*]	0.006
PT2	0.148[*]	0.035	0.156[*]	0.038	0.135[*]	0.033
PT3	0.129[*]	0.035	0.133	0.038	0.116	0.033
PT4	0.221[**]	0.038	0.242[**]	0.038	0.272[**]	0.036
CLS	0.289[**]	0.043	0.301[**]	0.048	0.276[**]	0.040
BIG	0.226[**]	0.027	0.236[**]	0.028	0.200[**]	0.025
Log PRC	−0.172[**]	0.032	−0.177[**]	0.030	−0.156[**]	0.025
WND	0.198[**]	0.032	0.201[**]	0.032	0.177[**]	0.029
NGT	−0.013	0.051	−0.014	0.052	−0.001	0.047
RAN	−0.002[**]	0.001	−0.002[**]	0.001	−0.002[**]	0.001
Constant	8.236[**]	0.496	8.725[**]	0.491	8.875[**]	0.428
Year dummies	Yes		Yes		Yes	
n observations	3175		3175		3175	
Log-likelihood	−3385		−3251		−32,593	

Note. Asterisks denote significance:
[*] Significant at 5%.
[**] Significant at 1%. Heteroskedasticity-consistent standard errors (std. error).

the Inverse Gaussian GLM and the variables' signals and statistical significance were very similar. The Inverse Gaussian GLM results are available by requesting authors.

As in Brazilian games the manager sets the ticket price and it remains fixed despite the attendance, we consider that price and quantity are not simultaneously determined, so there is no endogeneity.

Based on the results reported in Table 5, we see that all models presented similar results for the signal and statistical relevance of the variables affecting ticket consumption. Variables representing the economic environment, that is, the resident population and annual per capita income in the city in which the game occurred were statistically significant in explaining consumption. With respect to population, the impact was positive, as expected. These results confirms hypothesis 1. The negative

impact with respect to income means that tickets for the games can be considered inferior goods. This result is in line with those found by Bird (1982) and to that observed in Madalozzo and Villar (2009). It could be argued that this negative effect is associated with the existence of other forms of entertainment in those cities with a higher per capita income, although this assertion requires further empirical study.

Among the variables that indicate the quality of the game, the majority were statistically significant. The current situation of the teams based on the ranking, points won and goals scored in the last three games has statistically significant influence on demand. We can observe a positive influence of the ranking on the audience on game day; that is, the better the team's position in the tournament, the bigger the audience in the stadium. Only the offensive power of the home team, represented by points won
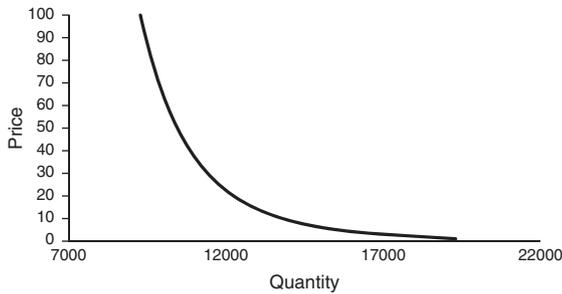
Fig. 3. Estimated demand curve.

and goals scored, has a significant and positive relationship with the number of tickets sold for such a game, thus confirming the hypothesis 2 that the higher the expected quality of the match, the greater the number of fans at the game.

As for the stage of the tournament, it should be noted that the demand increases as the championship progresses, that is, the final phase (stage 4) attracts more attendees than the earlier phases and the intermediate stages (stages 2 and 3) attract more attendees than the initial phase (stage 1). As expected, the fact that a game is a classic, that is, a rivalry between major teams with a strong fan base, has a positive effect on the demand for tickets, as indicated in Wooten (2015). Moreover, if the visiting team is one of the major league teams from Sao Paulo or Rio de Janeiro, the demand for tickets increases.

Variables representing the incentives that the public has to go to the stadium, such as the average ticket price, the day of the game and the amount of rainfall also play a significant role on demand, confirming hypothesis 3.

The average ticket price reflected a negative sign, thus indicating that the higher the price, the lower the demand and the price elasticity of −0.172 is similar to that found in Bird (1982), Nilon (2010) and Madalozzo and Villar (2009), what shows that the soccer demand is inelastic, according to our point of view that Brazilian soccer managers do not seek profit maximization. This elasticity means that, everything else held constant, an increase in prices leads to a decrease in lower proportion in quantity demanded, so that there is an increase in total revenue. In Fig. 3, we plot the estimated demand curve for a classic match between big national teams in a sunny weekend afternoon, with all others variables held constant in their average values. As one can see, if the price is set at its medium value (R$20.36), the revenue is about R$248,452, while a price of R$50.00 generates a revenue of R$522,783, more than 110% of increase.

The lowest attendance in the stadiums coincided with game days that had the most rainfall in the area of the game, reflecting the potential problems with transportation on rainy days. Finally, the years are significant variables in controlling the increase in attendance between 2004 and 2013, as can be seen in Fig. 2.

As mentioned previously we left 400 observations out of sample to evaluate the accuracy of forecasting models. The forecasts results were evaluated according to three criteria commonly used to compare them: the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). These measures are computed as follows:

Table 6
RMSE, MAE and MAPE to compare the performance of the forecasts in-sample and out-of-sample for the regression models.

Model	LRM	TOBIT	Gamma-GLM
RMSE			
In sample	9866	9548	**9429**
		(3.22)	**(4.43)**
Out sample	42,349	38,026	**30,064**
		(10.21)	**(29.01)**
MAE			
In sample	6707	6641	**6382**
		(0.99)	**(4.85)**
Out sample	10,938	9165	**7238**
		(16.21)	**(33.83)**
MAPE			
In sample	73.34	72.56	**61.65**
		(1.06)	**(15.94)**
Out sample	59.32	58.03	**50.62**
		(2.17)	**(14.66)**

Notes. The outperformed result for each line is shown in bold. Percentage improvement in the forecast of each model relative to the LRM (benchmark) in parentheses: LRM relative % = [1 − (MODEL/LRM)] × 100. Best result is in bold.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{k}(y_i - \hat{y}_i)^2}{k}}, \quad \text{MAE} = \frac{\sum_{i=1}^{k}|y_i - \hat{y}_i|}{k}$$

$$\text{and} \quad \text{MAPE} = \frac{100}{k}\sum_{i=1}^{k}\frac{|y_i - \hat{y}_i|}{|y_i|}, \tag{6}$$

where y_i and $\hat{y}_i$ are the true demand and the forecasted demand for the game i, respectively, and k is the number of forecasts. We are particularly interested in the RMSE because this measure should reflect the forecasted standard deviations of the estimated model.

Table 6 presents the RMSE, MAE and MAPE of all the estimated models for the data set for the in-sample ($k = 3175$) and out-of-sample observations ($k = 400$). As the LRM is commonly employed in the most varied situations, we chose it as the benchmark to evaluate the quality of forecasts. The LRM has upper values for the three estimated forecasting errors and, hence, shows inferior accuracy compared with the TOBIT and Gamma-GLM models both for the in and out-of-sample analysis.

The GLM with Gamma distribution presented the best forecasting performance when compared with the LRM with Normal distribution and TOBIT model, which was expected since such models employ variables with a distribution that allows for a better fit for the data skewness. By evaluating the in-sample forecasts, the Gamma-GLM's results outperform the LRM in more than 15% according to MAPE. For the out-of-sample forecasts the Gamma-GLM shows results 29% and 14% more accurate than those for the LRM using the RMSE and MAPE measures, respectively, indicating that Gamma-GLM was superior to the other estimated models to forecast the ticket consumption for Brazilian games.

Conclusion

The number of fans who attend the stadiums in Brazil is not considered satisfactory, especially when compared with European standards. This situation constitutes a problem in the development of Brazilian soccer as the revenue earned by a soccer club through ticket sales could represent its major financial resource.

The aims of this study were to demonstrate how ticket consumption is affected by several variables so that corrective actions can be adopted to increase both the presence of the public in stadiums and the steadiness of the attendance. The applied models included variables related to the economic environment, the product quality and the incentive that people have to attend a sporting event at a stadium. The econometric approach allows us to more accurately forecast the ticket consumption of Brazilian soccer games. The Gamma-GLM presented greater performance than the usual LRM, indicating that it is possible to obtain better forecasts for consumption using more general regression models. In the case of the TOBIT model, which takes into account that the observed consumption is censored by the capacity of the stadium, the results were very similar to the LRM due to the low percentage of censure (3.1%).

The three regression models used for the paying public presented similar results related to statistical relevance and signal variables. The analysis shows that all variables related to the economic environment, quality of the product and incentives people are given to attend a match at a stadium are important to explain the consumption of soccer games in Brazilian stadiums, except those indicating recent visiting team performance and whether the match occurred after 9:00 pm. These are very important results for team managers. Since they are not able to control economic variables, managers should focus on quality and incentive variables. As such, from a strategic point of view, three main conclusions can be drawn from the results.

First, teams have to increase incentives if they want to keep a loyal base of fans and managers should not overlook the effect of a weak infrastructure in the process, as the study was able to show that rainfall negatively influences the demand for tickets, and the lack of covered stadiums (in whole or in part) is the rule rather than the exception in Brazil. Fortunately, almost a dozen of new arenas with much better infrastructure were built in Brazil to host 2014 World Cup and managers should target long term rental agreements with the owners of these arenas, which are in many cases state-owned. Stadiums should feel like home for fans, helping teams acquire a stronger identity and eventually leading to more ticket sales. Financial distress of Brazilian teams should not be considered a deal breaker as a public-private partnership may be implemented. Many states are probably eager to get rid of maintenance costs (or at least part of it) associated with the already built arenas.

The second strategy is linked to the overall performance of both the challenging team and the home team. These factors, notably the latter, have a tremendous influence on the demand for tickets and are of paramount importance to club managers with respect to taking corrective action so as to stimulate the influx of fans to stadiums. To mitigate the negative effect on the demand (and therefore on the revenue earned) of a team's (and its sub-par) performance, it is essential that a significant portion of the tickets be sold before the beginning of the championships as is successfully done in Europe where clubs offer ticket packages with great advantages to the fans. Teams should successfully appeal to dedicated fans.

Finally, our results suggest that Brazilian soccer managers do not seek profit maximization. Whether this is a consequence of pressure from well organized groups like football firms or any other factor is yet to be demonstrated by other studies. But this approach to set the price of tickets leads to a significant decrease on the expected revenue, hampering the ability of the soccer club to provide fans with those incentives discussed above. As such, Brazilian managers should change the basis of setting ticket prices.

The analysis is not exhaustive and there is room for further studies that, for example, will demonstrate the effect of the habit of going to the stadium or, additionally, will explain the negative influence of the per capita income of the city on the ticket consumption of soccer games. Nevertheless, the results obtained are robust and can be used as subsidies for the strategic planning of soccer associations in Brazil.

Conflicts of interest

The authors declare no conflicts of interest.

References

Allan, G., & Roy, G. (2008). Does television crowd out spectators? New evidence from the Scottish premier league. *Journal of Sports Economics*, *9*(6), 592–605. http://dx.doi.org/10.1177/1527002508321458

Alonso, A. D., & O'Shea, M. (2013). The links between reasons for game attendance of a new professional sports league and revenue management: An exploratory study. *International Journal of Revenue Management*, *7*(1), 56–74. http://dx.doi.org/10.1504/IJRM. 2013.053359

Baimbridge, M., Cameron, S., & Dawson, P. (1996). Satellite television and the demand for football: A whole new ball game? *Scottish Journal of Political Economy*, *43*(3), 317–333. http://dx.doi.org/10.1111/j.1467-9485.1996.tb00848.x

Bird, P. J. W. N. (1982). The demand for league football. *Applied Economics*, *14*(6), 637–649. http://dx.doi.org/10.1080/00036848200000038

Borland, J., & Lye, J. (1992). Attendance at Australian Rules football: A panel study. *Applied Economics*, *24*(9), 1053–1058. http://dx.doi.org/10.1080/00036849200000083

Calster, B. V., Smits, T., & Huffel, S. V. (2008). The curse of scoreless draws in soccer: The relationship with a team's offensive, defensive, and overall performance. *Journal of Quantitative Analysis in Sports*, *4*(1.) http://dx.doi.org/10.2202/1559-0410.1089

Cox, A. (2012). Live broadcasting, gate revenue, and football club performance: Some evidence. *International Journal of the Economics of Business*, *19*(1), 75–98. http://dx.doi.org/10.1080/13571516.2012.643668

Falter, J.-M., & Pérignon, C. (2000). Demand for football and intramatch winning probability: An essay on the glorious uncertainty of sports. *Applied Economics*, *32*(13), 1757–1765. http://dx.doi.org/10.1080/000368400421101

Falter, J.-M., Pérignon, C., & Vercruysse, O. (2008). Impact of overwhelming joy on consumer demand: The case of a soccer world-cup victory. *Journal of Sports Economics*, *9*(1), 20–42. http://dx.doi.org/10.1177/1527002506296548

García, J., & Rodríguez, P. (2002). The determinants of football match attendance revisited: Empirical evidence from the Spanish football

league. *Journal of Sports Economics*, *3*(1), 18–38. http://dx.doi.org/10.1177/152700250200300103

Giovannetti, B., Rocha, B. P., Sanches, F. M., & Silva, J. C. D. (2006). Medindo a fidelidade das torcidas brasileiras: Uma análise econômica no futebol. *Revista Brasileira De Economia*, *60*(4), 389–406. http://dx.doi.org/10.1590/s0034-71402006000400004

Greene, W. H. (2003). *Econometric analysis* (5th ed.). New Jersey: Prentice Hall.

Gujarati, D. N. (2004). *Basic econometrics* (4th ed.). New York: McGraw-Hill/Irwin.

Jennett, N. (1984). Attendances, uncertainty of otucome and policy in Scottish league football. *Scottish Journal of Political Economy*, *31*(2), 176–198. http://dx.doi.org/10.1111/j.1467-9485.1984.tb00472.x

Knowles, G., Sherony, K., & Haupert, M. (1992). The demand for major league baseball: A test of the uncertainty of outcome hypothesis. *American Economist*, *36*(2), 72–80. Retrieved from http://www.jstor.org/stable/25603930

Kotler, P. (1975). *Marketing for non-profit organizations*. New Jersey: Prentice Hall.

Madalozzo, R., & Villar, R. (2009). Brazilian football: What brings fans to the game? *Journal of Sports Economics*, *10*(6), 639–650. http://dx.doi.org/10.1177/1527002509335572

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

Melo, L. M. (2007). Brazilian football: Technical success and economic failure? In R. D. Miller, & L. Crolley (Eds.), *Football in the Americas*. Institute of Latin American Studies.

Mentzer, J. T., & Moon, M. A. (2005). *Sales forecasting management: A demand management approach* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Nilon, J. (2010). *An inelastic phenomenon: Analyzing the elasticity of demand for soccer in England per geographical region (creative distraction 2010, 14th edition)*. pp. 41–57. Retrieved from University of Queensland Economics Society website: http://www.uqes.com.au/system/files/ESA%20Creative%20Distraction%202010%20Edition%2014_1.pdf

Park, K. W., Lee, S., & Miller, P. (2012). Ticket pricing per team: The case of major league baseball (MLB). In *Allied academies international conference: Proceedings of the academy of economics and economic education* (p. 15). Retrieved from http://www.alliedacademies.org/pdfs/AEEE_Proceedings_15%282%29_Fall_2012.pdf

Santos, R. B., Bazanini, R., & Ferreira, A. A. (2014). Entrepreneurship in the society of spectacle: Soccer management in a globalized world. *Global Journal of Management and Business Research: A Administration and Management*, *14*(9), 57–68. Retrieved from http://journalofbusiness.org/index.php/GJMBR/article/view/1477/1384

Simmons, R. (1996). The demand for English league football: A club-level analysis. *Applied Economics*, *28*(2), 139–155. http://dx.doi.org/10.1080/000368496328777

Sloane, P. (1997). Editorial: The economics of sport: An overview. *Economic Affairs*, *17*(3), 2–6. http://dx.doi.org/10.1111/1468-0270.00034

Smith, E. E., & Groetzinger, J. D. (2010). Do fans matter? The effect of attendance on the outcomes of major league baseball games. *Journal of Quantitative Analysis in Sports*, *6*(1) http://dx.doi.org/10.2202/1559-0410.1192

Souza, F. A. P. (2004). *Um Estudo Sobre A Demanda Por Jogos De Futebol Nos Estádios Brasileiros* (Master's Degree dissertation). São Paulo, SP, Brazil: Faculdade de Economia, Administração e Contabilidade – Universidade de São Paulo.

Szymanski, S. (2001). Income inequality, competitive balance and the attractiveness of team sports: Some evidence and a natural experiment from English soccer. *Economic Journal*, *111*(469), 69–84. http://dx.doi.org/10.1111/1468-0297.00599

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, *26*(1), 24–36. http://dx.doi.org/10.2307/1907382

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge: The MIT Press.

Wooten, J. J. (2015, March). *Rivalries in US professional soccer*. http://dx.doi.org/10.2139/ssrn.2394631. Retrieved from http://ssrn.com/abstract=2394631