XVIII SEMEAD
Seminários em Administração
novembro de 2015
ISSN 2177-3866

DEMAND FORECAST FOR BRAZILIAN CHAMPIONSHIP GAMES USING GENERALIZED LINEAR REGRESSION MODEL

MAURICIO MESQUITA BORTOLUZZO
Universidade Presbiteriana Mackenzie
mauriciomesquita@hotmail.com

ADRIANA BRUSCATO BORTOLUZZO
Insper Instituto de Ensino e Pesquisa
adrianab@insper.edu.br

SÉRGIO JURANDYR MACHADO
Secretaria de Política Econômica
juramachado@globo.com

DEMAND FORECAST FOR BRAZILIAN CHAMPIONSHIP GAMES USING GENERALIZED LINEAR REGRESSION MODEL

Area: Marketing

Abstract

For the efficient of sales and marketing management of athletic clubs, it is crucial that there be a way to appropriately estimate the level of demand for the sporting events. More precise estimates allow for an appropriate financial and operational plan, which results in better club performance and a higher quality of service delivered to the fans. The focus of this study is to analyze and to forecast the demand 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 and Poisson generalized linear model. The models included explanatory variables related to the economic environment, product quality and monetary and non-monetary incentives that people have to go to the stadium, and we showed that most of these variables are statistically significant to explain the amount of fans that go to the stadiums. We used different measures of accuracy to evaluate the performance of demand forecasts and concluded that generalized linear models presented better results compared to the benchmark. Among these last models, the best predictions were obtained using the Gamma distribution, which allowed a better fit of the positive skewness of demand for tickets in the Brazilian Championship.

Keywords: sports management; generalized linear regression model; demand; soccer.

1 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 over 100 years old, have many organizational problems and, in general, are overwhelmed by 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, 6.7 million fans attended soccer games in Brazil's stadiums during the 2009 season – according to Brazilian Soccer Confederation (CBF) –, an amount that could be considered minimal when compared to the 27.1 million fans that attended games in British stadiums during the same season. 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, and can never be predicted perfectly. Yet, the marketer or the administrator must plan and take decisions using whatever constitutes the best estimate about future developments. Without a proper demand forecast the marketing executive cannot determine the type of marketing program to use in order to attain the desired sales and

marketing objectives. Therefore, evaluating the demand potential and preparing a demand forecast is an important function of sales and marketing managers. Mentzer and Moon (2004) defines 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 these sporting events. Such analyses will prove instrumental for the Brazilian clubs as these clubs seek to increase attendance, and will thus contribute to a more efficient and professional management of the sport. The focus of this study, therefore, is to examine the demand for soccer stadiums in Brazil through an analysis of the paying public at the matches of the A series of the Brazilian Championship (*Campeonato Brasileiro*) between 2004 and 2009.

First, it is necessary to determine which variables influence the demand for games using regression models. Four models were fitted: an 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 two generalized linear models (GLM) with Gamma distribution to better adjust the positive skewness of the attendance distribution and with Poisson distribution to count data. For this study, it is determined that the TOBIT and the GLM models are econometrically more appropriate than other regression models as the literature on the topic related to the Brazilian case (Madalozzo & Villar, 2009; Melo, 2007, Giovannetti *et al.*, 2006; Souza, 2004) 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 decision-making process within the ambit of Brazilian soccer clubs, allowing better in-sample and out-sample demand forecasts.

This study contributes to the sports management literature in two ways. First, our framework integrates different types of exploratory variables, which include those related to the economic environment, product quality and monetary and non-monetary incentives that people have to go to the stadium, to explain the demand for soccer games in Brazil. For managers, our paper elucidates the relevant measures to become the focal point of the clubs effort to leverage its revenue from the games. Second, we contribute to the literature by using the generalized regression models that are able to produce more reliable forecasts of demand for soccer tickets.

The remainder of this paper is divided into the following sections: Section 2 presents the literature review and description of variables incorporated in the model based on the theoretical framework, Section 3 briefly describes the methodology, Section 4 presents the interpretation of results, and Section 5 presents conclusions and suggestions for future research.

2 Literature Review and Description of Study Variables

A merely superficial analysis of the management process of Brazilian soccer 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 sale of broadcasting 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 to clubs to study how these variables affects the demand for championship games, so they are able to make reliably predictions of revenue from soccer tickets.

One of the crucial issues of effective management is embodied in the estimation of the levels of demand for those sporting events in which a particular club will participate throughout the year. The more precise the estimate, the better able management is to plan appropriately for the organization's financial and operational needs. A better plan, in turn, tends to result in a better performance of the club on the field and a higher quality of service delivery to the fans. Together, these two factors lead to a greater number of fans in the stadium and, consequently, increased revenue generated by the event for the organization, thus forming a profitable cycle.

As pointed out by Smith and Groetzinger (2010), "it is possible it would be profitable for clubs to drop prices for the purpose of increasing the chance of victory, as the revenue associated with an additional win could outweigh loss ticket revenue". Even if that is the case, it is essential to understand consumers' demand for soccer games.

Garcia and Rodriguez (2002) specified a demand 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 demand 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 use this last specification.

The group of economic environment variables, defined here as structural variables, affects demand in that a consumption of the good (the purchase of tickets) affects the consumer's budget and is limited by available income. Therefore, two variables were incorporated into the model: per capita income in the city where the game occurs (PCI) and the population of the city (POP). Data on these variables were obtained from the website of the Brazilian Institute of Geography and Statistics (IBGE).

Income elasticity of demand measures the sensitivity of the demand for a good related to changes in consumer income, ceteris paribus. A negative income elasticity of demand is associated with inferior goods while a positive 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 tickets to soccer matches are an inferior good. Madalozzo (2009) found a negative relationship between the average per capita income and the demand for tickets in Brazilian games.

Thus the following hypotheses can be made:

H1a: The average per capita income of the city negatively influences the demand for soccer tickets in the Brazilian Championship.

H1b: The size of city's population impacts positively the demand for soccer tickets in the Brazilian Championship.

The second group includes those variables related to product quality, that is, the quality of the soccer team that sells its ticket 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 demand for 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 the demand for tickets. 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 derived 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 the scoring of a goal is the part of the game the fans most enjoy and appreciate. Games with more goals tend to be viewed as more entertaining games as pointed out by Van Calster, Smits and Van Huffel (2008).

The importance of the match in the league can also affect demand. It is reasonable to assume that a game with a higher relative value will attract more attendees. Thus, 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 according 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 is an increase in the demand for the game. Aiming to represent the effect of rivalry between teams, an indicator 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.

A variable to represent the presence of a large team from São 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 São Paulo or Rio de Janeiro faces the Sao Paulo, Corinthians, Palmeiras, Santos, Flamengo, Fluminense, Botafogo or Vasco teams, and assumes a value of zero for other cases. It is expected that the presence of big clubs from these two states will increase the demand for tickets for the game in question.

Thus, the following hypotheses are made about the relationship between the demand for tickets and the quality of the teams in Brazilian Championship:

H2a: The better the rating of the home team (or opponent), the greater the demand for tickets.

H2b: The higher the number of points won by the home team (or the opposing team) in the last three games, the greater the demand for tickets.

H2c: The greater the number of goals scored by home team (or the opposing team) in the last three games, the greater the demand for tickets.

H2d: More advanced stages of the championship have higher demand compared to the earlier stages of the championship.

H2e: Classic matches have greater demand for tickets.

H2f: Matches with participation of big teams from São Paulo and Rio de Janeiro have greater demand for tickets in the Brazilian Championship.

The third group of variables represents the incentives that fans have to go to the stadium. There is a significant monetary incentive associated with the ticket price. Accordingly, was included a variable representing the average ticket price (PRC), obtained by dividing the income earned in the game by the number of paying attendees. When price elasticity of the demand of a good is elastic, lowering its price causes an increase in revenues, and when its demand is inelastic revenues fall. Thus, the price elasticity is related to product's marginal revenue, and one profit-maximizing firm has to sell its product in the elastic portion of its demand curve, where marginal revenue is positive, when its marginal costs are positive, or sell its product in the unit elasticity portion of its demand curve, when its marginal costs are zero. We can conclude this should be a good approximation to the case of the marginal cost of soccer match tickets.

Several studies found evidence of inelastic demand for tickets across sports and countries: Jennett (1984) in the Scottish soccer league, Borland and Lye (1992) in the Australian Rules soccer, Garcia and Rodriguez (2002) in the Spanish soccer league, and Madalozzo (2009) for Brazilian soccer league.

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 are played have a significant influence on the demand. 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 21:00 (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. Factors such as extreme heat or cold, or rain, can be major factors as fans decide whether to go to the stadium to watch the game. To imprint this effect into the study, we included a variable that represents the accumulated precipitation in mm on the day and in the city where the match occurred (RAN) and a variable that expresses the season during which the match was played. It can be expected that the more total rainfall recorded on game day, the lower the incentive for fans to attend game depending on i) the deficient infrastructure of the stadiums in Brazil (for example, not all seats are covered or protected 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.

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 at the site 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).

The suggested hypotheses for the factors that represent the incentives that fans have to go to the stadiums are:

H3a: The ticket price negatively influences the demand for tickets.

H3b: The greater the non-monetary incentives, the greater the demand for tickets

Because the matches span the years 2004 to 2009, we will include variables indicating the year of the matches (Y05, Y06, Y07, Y08 and Y09) to analyze the demand over evaluated time and to monitor the possible effect of time in the demand for tickets.

Table 1 presents a summary describing the variables that were added to the models and the expected effect of each of them on the demand.

Table 1
List of Explanatory Variables and the Expected Effect

	Variable		Description	Expected effect
structural variables	PCI	H1a	Annual per capita income in the city where the game takes place (in R$)	-
	POP	H1b	City population where the game occurs	+
variables related to product quality (soccer specific variables)	CLH	H2a	Classification of the home team	-
	CLV	H2a	Classification of the visiting team	-
	PGH	H2b	Points won by the home team in the past 3 games	+
	PGV	H2b	Points won by the visiting team in the past 3 games	+
	GLH	H2c	Goals scored by the home team in the past 3 games	+
	GLV	H2c	Goals scored by the visiting team in the past 3 games	+
	PT2	H2d	Value 1 if the match occurs on the 2nd stage of the championship	+
	PT3	H2d	Value 1 if the match occurs on the 3rd stage of the championship	+
	PT4	H2d	Value 1 if the match occurs on the 4th stage of the championship	+
	CLS	H2e	Value 1 if the match is considered a classic	+
	BIG	H2f	Value 1 if the game has some great team from SP or RJ	+
Incentives variables	PRC	H3a	Average ticket price	-
	WND	H3b	Value 1 if the match occurred on the weekend	+
	NGT	H3b	Value 1 if the match occurred after 21:00	+
	RAN	H3b	Rainfall (in mm)	-

Finally, one explanatory variable related to the television broadcasting of the soccer match must be factored into the equation, as Grant and Graeme (2008) have warned that such transmissions negatively impact 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 there was total capacity sell out for the stadium. Although there is a relationship of competition between ticket sales and the sale of pay-per-view packages, we do not consider it a problem for our study due to the low number of TV subscribers by 2009. However this should be a problem increasingly relevant as the TV subscriptions has grown exponentially from 2009 onwards. From late 2010 to late 2011 for example there was a 30.5% increase in the number of pay TV subscribers, according to the Brazilian National Agency of Telecommunication (ANATEL).

3 Methodology

This study has two main objectives. First, we want to evaluate how the variables related to the economic environment, product quality, and incentives to go to the stadium influence the demand for Brazilian Championship matches. Our second goal is to make accurate predictions for the demand. The econometric approach used thus far by studies on the Brazilian context is not ideal as it neglects the fact that there exists a restriction due to the capacity of stadiums and that the demand has a positive skewness. Aiming to make better forecasts of demand, we decided to work with more general regression models than the usual multiple linear regression model with normal distribution for the errors. We chose the TOBIT model, that takes into account that the observed demand is censored by capacity the stadium, the generalized linear regression model (GLM) with Gamma distribution for the response variable in order to adjust the positive skewness of demand, and the Poisson regression model that is used to count data since the demand is the amount of paying individuals that go to the stadium.

3.1 TOBIT model

A censored sample is one in which the information returned is not available for all observations, that is, in the case of data on public demand in stadiums, there is the real demand only for those matches in the stadiums that did not exceed the maximum capacity. Thus, the real demand for tickets should not be estimated based on matches in which the stadium is filled. 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). Thus, marginal effects calculated based on the OLS estimators would not be suitable for the real situation and, therefore, would lead to incorrect interpretations of the impact of each of the variables on demand.

The appropriate regression model for these cases is the TOBIT, as it allows for working with censored variables; that is, where the real value is not noticed. In this work, the actual demand for tickets has not been observed in games where tickets were sold out; that is, the demand was censored on the right and the observed value corresponds to the limit of tickets on sale (stadium sold out).

The TOBIT model was proposed by Tobin (1958) and is named in honor of its creator. The TOBIT model specification is similar to the multiple linear regression model and is given by

$$Y = X\beta + \varepsilon ,$$

where Y is the value of demand, that is, it corresponds to the value of actual demand if the tickets are not sold out, or it corresponds to the value of the stadium capacity if the tickets are sold out; X represents the matrix of explanatory variables, and ε is the random error with zero mean and variance σ^2.

In our case, Y is the paying public (PUB) who attended the soccer matches of the Brazilian Championship from 2004 to 2009. Intuitively, we can see that this value cannot assume a negative value. The threshold demand value represents the maximum attendance of a game as defined by the capacity of the stadium where the match is played. Truncation occurs in games where the demand for tickets is greater than the number of people the stadium can support; such that demand is designated as equal to capacity of the stadium.

The TOBIT model is estimated using the method of maximum likelihood estimation. However, due to the censure observed, the process involves a weighting of uncensored with censored variables. The goal is to eliminate the bias of estimators and adapt the method for latent models. It is noteworthy that the estimated coefficients for the TOBIT model should not be used directly to assess the impact of a change of a unit in the value of the regressor on the dependent variable. The calculation of the marginal effects of explanatory

variables on demand can be easily obtained, and depends on the amount of censored data (Greene, 2003, p. 765). According to Greene (2003), the marginal effects of the TOBIT model will vary significantly from the marginal effects via OLS, due to the higher percentage of censored data.

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

3.2 Generalized Linear Model

The idea of GLM is to expand the range of options for the response variable 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. Thus, 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

$$E(Y_i) = \mu_i = g^{-1}(X_i\beta),$$

where $X_i\beta$ is the linear predictor and g is a monotonic and differentiable function, called link function.

In this work, the response variable is the amount of people paying to go to a soccer game, and then we selected the Poisson distribution to represent these count data (Goddard, 2005). Regression models for forecasting goals and match results in association soccer.. So, the regression Poisson model with log link function can be written as

$$E(Y_i) = \mu_i = \exp(X_i\beta) \quad \text{and} \quad Var(Y_i) = \mu_i .$$

Note that the variance of the dependent variable is equal to the mean and there may be heterogeneity among observations.

The Gamma distribution is a positive skewed distribution and we choose the log link function, that is the mean and variance depend on the exploratory variables through

$$E(Y_i) = \mu_i = \exp(X_i\beta) \quad \text{and} \quad Var(Y_i) = \mu_i^2 .$$

We used the maximum likelihood method to estimate the parameters of Gamma-GLM and Poisson with the Quadratic Hill Climbing optimization algorithm. More details about the GLM, its estimation and its asymptotic properties of estimators can be found in McCullagh and Nelder (1989).

4 Results

As mentioned earlier, the dependent variable in the models to be analyzed represents the demand for tickets at the stadiums (PUB), that is, the paying public for the games analyzed. These data are obtained from summaries and bordereau at the site of the Brazilian Soccer Confederation (CBF) from 2004 to 2009. It is important to note that obtain these data is quite complicated because the data from each of the games are not available and it is necessary to collect data from each individual game. In total, there were 2,481 games, and only 31 (1.25%) of which had maximum capacity and were regarded as censored in the TOBIT model.

Figure 1 shows the graph of the evolution of the paying public and the capacity of the stadiums in the Brazilian Championship from 2004 to 2009. The average occupancy rate of stadiums was 32% of full capacity during this period, a low percentage, especially when compared to European standards. 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. 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 2009. As there was this increase in audience over the assessed time, we will include indicators of the year in the model to control for this effect in time.

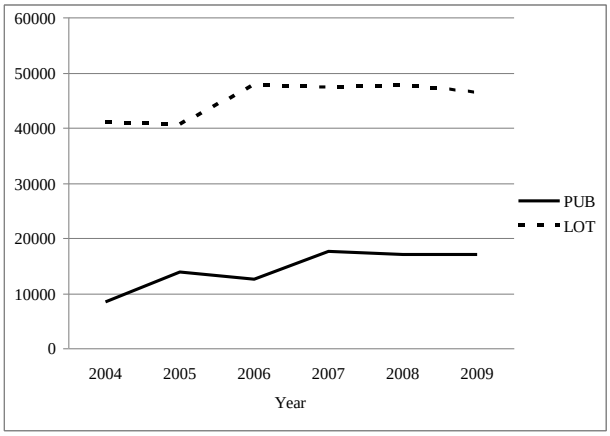

Figure 1. Evolution of the Paying Public and the Capacity of Stadiums in the Brazilian Championship from 2004 to 2009.

In the descriptive analysis of the categorical variables, we see that 8.5% of the games were considered classic, 36.7% of the games include the major teams from São Paulo or Rio de Janeiro as visitors, 74.4% of the games occurred on weekends and 17.5% of the games took place after 21:00. It was made the correlation matrix of the explanatory variables and was not found a high correlation between variables, so there is no strong multicollinearity.

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 13,937 fans per game with a high dispersion of 11,509. The average capacity of the stadiums in the Brazilian Championship games during the specified period was 44,665, more than three times the paying public. There is great variability in these values with a minimum capacity of 11,400 and a maximum of 95,000 people. It is important to note the existence of positive skewness in the paying public, which guided the choice of the generalized linear model (GLM) with gamma distribution for a better data fit.

Table 2
Descriptive Analysis of Numerical Variables

Variable	Average	Standard deviation	Minimum	Maximum	Skewness
PUB	13,937	11,509	147	87,895	1.99
LOT	44,665	24,276	11,400	95,000	0.69
PCI	22,773	11,833	6,895	100,806	3.25
POP	4,603,280	6,226,333	87,800	19,223,897	1.72
CLH	10.85	6.26	0	24	0.05
CLV	10.43	6.27	0	24	0.12

PGH	3.88	2.29	0	10	0.27
PGV	4.24	2.31	0	12	0.15
GLH	4.09	2.16	0	13	0.48
GLV	4.32	2.20	0	14	0.46
PRC	12.54	7.16	0	148	5.51
RAN	3.49	9.63	0	86	4.23

The model estimation used 2,281 observations and we left 200 observations outside the model to assess the out of sample forecasts. The sizes of our in and out of sample estimations were defined to guarantee the robust estimation of the models according to the number of explanatory variables (24 variables). We then compared the accuracy of the forecasts of the in-sample model using the in-sample observations and the out of sample observations.

We use four specifications for the regression model that has the natural logarithm of paying public in the games of the Brazilian Championship as the response variable: (1) usual linear regression model (LRM) with normally distributed errors, (2) TOBIT model truncated by the maximum capacity of the stadiums (LOT), (3) generalized linear model (GLM) using a Gamma distribution to better fit the degree of positive skewness of the response variable and (4) a model for count data using the Poisson distribution for the paying public. The estimated model coefficients and their standard errors are in Table 3. For each model an analysis of residuals was conducted to verify the assumptions about the errors and the necessary corrections have been implemented.

Table 3

Estimated Regression Models for Public Demand in the Brazilian Championship Games from 2004 to 2009

	LRM		TOBIT		Gamma-GLM		Poisson-GLM	
Variable	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Log RPC	-0.4995**	0.0411	-0.5080**	0.0388	-0.0578**	0.0049	-0.3528**	0.0360
Log POP	0.1501**	0.0123	0.1463**	0.0117	0.0167**	0.0014	0.1384**	0.0108
CLH	-0.0256**	0.0026	-0.0256**	0.0026	-0.0028**	0.0003	-0.0247**	0.0025
CLV	-0.0088**	0.0025	-0.0089**	0.0025	-0.0009**	0.0003	-0.0071**	0.0025
PGH	0.0436**	0.0076	0.0438**	0.0078	0.0048**	0.0008	0.0333**	0.0074
PGV	-0.0062	0.0079	-0.0056	0.0080	-0.0006	0.0009	-0.0113	0.0079
GLH	0.0345**	0.0079	0.0351**	0.0079	0.0037**	0.0009	0.0324**	0.0076
GLV	0.0172*	0.0078	0.0168*	0.0078	0.0018*	0.0009	0.0223**	0.0079
PT2	0.1360*	0.0623	0.1400*	0.065	0.0146*	0.0068	0.1425*	0.0596
PT3	0.0961*	0.0679	0.0995	0.0722	0.0099	0.0074	0.1009	0.0646
PT4	0.1474	0.0921	0.1526	0.0935	0.0146	0.0101	0.2041*	0.0883
CLS	0.4313**	0.0511	0.4460**	0.0548	0.0474**	0.0054	0.3116**	0.0502
BIG	0.2541**	0.0312	0.2566**	0.0319	0.0282**	0.0034	0.1891**	0.0312
Log PRC	-0.2580**	0.0400	-0.2577**	0.0354	-0.0284**	0.0045	-0.2186**	0.0316

WND	0.2441**	0.0595	0.2481**	0.0592	0.0263**	0.0065	0.2393**	0.0567
NGT	0.0557	0.0686	0.0584	0.0679	0.0046	0.0075	0.0957	0.0677
RAN	-0.0051**	0.0018	-0.0051**	0.0016	-0.0006**	0.0002	-0.0041*	0.0017
Y05	0.3712**	0.0451	0.3726**	0.0457	0.0417**	0.0051	0.3266**	0.0412
Y06	0.1680**	0.0484	0.1679**	0.0477	0.0184**	0.0055	0.2233**	0.0466
Y07	0.5311**	0.0493	0.5345**	0.0501	0.0573**	0.0054	0.5812**	0.0477
Y08	0.4981**	0.0520	0.4962**	0.0496	0.0539**	0.0058	0.5524**	0.0468
Y09	0.4423**	0.0635	0.4440**	0.0631	0.0480**	0.0070	0.5211**	0.0638
Constant	11.9952**	0.4740	12.1271**	0.4719	2.5522**	0.0542	10.7600**	0.4293

Note. Asterisks denote significance: * Significant at 5%, and ** Significant at 1%. Heteroskedasticity-consistent standard errors.

Based on the results reported in Table 3, we see that all models presented similar results for the signal and statistical relevance of the variables affecting demand. 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 demand. With respect to population, the impact was positive, as expected (H1b). The negative impact with respect to income means that tickets for the games can be considered inferior goods (H1a). 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 (H2a), points won (H2b) and goals scored (H2c) in the last three games has statistically significant influence on demand. We can observe a negative influence of the ranking on the audience on game day; that is, the worse the team's position in the tournament, the smaller the audience in the stadium. Only the offensive power of the home team, represented by points won and goals scored, has a significant and positive relationship with the number of tickets sold for such a game, thus confirming the assumption that the greater the number of goals scored, the greater the number of fans at the game.

As for the stage of the tournament (H2d), it should be noted that the demand increases with the advance of the championship, that is, the final phase (stage 4) attracts more attendees than the earlier phases and the intermediate stages (steps 2 and 3) attract more attendees than the initial phase (stage 1). As expected, the fact that a game is a classic (H2e), that is, a rivalry between major teams with a strong fan base, has a positive effect on the demand for tickets. Moreover, if the visiting team is one of the great teams from São Paulo or Rio de Janeiro, the demand for tickets increases (H2f).

Variables representing the incentives that the public has to go to the stadium, such as the average ticket price (H3a), the day and the time of the game and the amount of rainfall (H3b) also play a significant role on demand. Only the seasons appear to have no effect on attendance.

The average ticket price reflected a negative sign, thus indicating that the higher the price, the lower the demand and the price elasticity of -0.26% is similar to that found in Bird (1982) showing that the soccer demand is inelastic. This 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.

Games that occur on weekends or after 21:00 have greater attendance than those held on weekdays before 21:00. 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 2009, as can be seen in Figure 1.

As mentioned previously we left 200 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:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{k}(y_i - \hat{y}_i)^2}{k}}, \quad \text{MAE} = \frac{\sum_{i=1}^{k}|y_i - \hat{y}_i|}{k} \text{ and } \text{MAPE} = \frac{100}{k}\sum_{i=1}^{k}\frac{|y_i - \hat{y}_i|}{|y_i|},$$

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 4 presents the RMSE, MAE and MAPE of all the estimated models for the data set for the in sample (k = 2,281) and out of sample observations (k = 200). As the LRM is commonly employed the most varied situations, we chose it as a benchmark to evaluate the quality of forecasts. The LRM has upper values for the three estimated forecasting errors and, hence, shows superior accuracy compared with all the other estimated models for the in and out of sample analysis.

Table 4

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	Poisson
RMSE				
in sample	9452	9448	9168	**8952**
		(0.05)	(3.01)	**(5.29)**
out sample	14146	14026	**11007**	12323
		(0.85)	**(22.19)**	(12.89)
MAE				
in sample	6345	6341	**6032**	6107
		(0.07)	**(4.94)**	(3.76)
out sample	9174	9165	**7047**	8087
		(0.10)	**(23.19)**	(11.85)
MAPE				
in sample	67.88	67.56	**59.34**	67.49
		(0.48)	**(12.59)**	(0.58)
out sample	70.75	70.82	**43.02**	62.06
		(0.05)	**(39.21)**	(12.29)

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.

The GLM with Gamma distribution and the Poisson model presented better forecasting performance when compared with the LRM with Normal distribution and TOBIT model, which was expected since such models employ variables with distribution that allows a better fit for the data skewness. By evaluating the in-sample forecasts, the Poisson regression model presented better results based on the RMSE while the Gamma-GLM showed better performance using the MAE and the MAPE measures. In the case of MAPE, the results of Gamma-GLM outperform the LRM in more than 12%. Out of sample best forecasts were obtained using the Gamma-GLM, which results were 22% and 39% 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 demand.

5 Conclusion

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

The aims of this study were to demonstrate how and to what extent the demand for tickets is affected by several variables so that corrective actions can be adopted to increase the presence of the public in stadiums and also choose better regression models to make accurate forecasts to the public paying demand in Brazilian Championship. The applied models included variables related to the economic environment, the product quality and the incentive that people have to go to the stadium.

The four regression models used for the paying public, that is the linear regression model with normally distributed errors, the TOBIT model, the generalized linear model using a Gamma distribution and the Poisson regression model, 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 have to go to the stadium are important to explain the demand for soccer games in Brazilian stadiums, except those indicating the season and the visiting team performance affect the demand for tickets.

The econometric approach used allows us to more accurately forecast the demand for tickets on Brazilian soccer games. The Gamma-GLM and the Poisson model presented greater performance than the usual LRM, indicating that it is possible to obtain better forecasts for demand using more general regression models. In the case of TOBIT model, which takes into account that the observed demand is censored by capacity the stadium, the results were very similar to the LRM due to the low percentage of censure (1.25%). The forecasts calculated here are of great importance for the managers of teams involved in competition to the extent that they explicitly and more accurately reflect the relative influence of explanatory variables.

The study further found that the degree to which a specific game influences the championship game plays a major role in explaining the demand for tickets. We estimated the relevance of variables related to the positions of the challenging teams and found the influence of these variables on the number of fans in the stadiums was significant. However, the most important conclusion with respect to variables is associated with product quality. The recent performance of the home team has a tremendous influence in the demand for tickets and is a determining factor when a fan decides whether to go to the stadium.

Together, these two findings 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 sub-par performance, especially with regard to recent 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 for the fans.

Much has been said about the need for an organized and invariant calendar that facilitates the advance sale of ticket packages. However, 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 nonexistence of covered stadiums (in whole or in part) is the rule rather than the exception in Brazil.

This study examined some of the factors that could affect the demand for tickets for the Brazilian Championship based on the foremost models described in literature regarding the subject. 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 per capita income of the city on the demand for soccer tickets. Nevertheless, the results obtained are robust and can be used as subsidies for the strategic planning of soccer associations in Brazil.

References

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*, 317-333.

Borland, J., & Lye, J. (1992). Attendance at Australian Rules football: A panel study. Applied Economics, 24, 1053-1058.

Bird, P. J. (1982). The demand for league football. *Applied economics, 14(6)*, 637-649.

Deserpa, A. C. (1994). To err is rational: A theory of excess demand for tickets. *Managerial and Decision Economics, 15(5)*, 511–518.

Falter, J. M., & Pérignon, C. (2000). Demand for football intra-match winning probability: an essay on the glorious uncertainty of sports. *Applied Economics, 32*, 1757-1765.

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*, 20-42.

Garcia, J., & Rodriguez, P. (2002). The determinants of football match attendance revisited: Empirical evidence from Spanish football league. *Journal of Sports Economics, 3*, 18-38.

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*, 389-406.

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting, 21*, 331-340.

Grant, A., & Greame, R. (2008). Does television crowd out spectators? New evidence from the Scottish premier league. *Journal of Sports Economics, 9*, 592-605.

Greene, W. H. (2003). *Econometric Analysis*. New Jersey: Prentice Hall.

Gujarati, D. N. (2004). *Basic Econometrics*. New York: McGraw-Hill/Irwin.

Jennett, N. (1984). Attendances, Uncertainty of Outcome and Policy in Scottish League Football. *Scottish Journal of Political Economy, 31*, 176–198.

Knowles, G., Sherony, K., & Haupert, M. (1992). The demand for major league baseball: a test of uncertainty of outcome hypothesis. *The American Economist, 36*, 72-80.

Madalozzo, R., & Villar, R. B. (2009). Brazilian football: what brings fans to the game*? Journal of Sports Economics, 10*, 639-650.

McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models*. Boca Raton: Chapman and Hall/CRC.

Melo, L. M. (2007). Brazilian football: technical success and economic failure. Miller, R. D. and Crowley, L. eds. Football in the Americas, Institute of Latin American Studies.

Mentzer, J. T., & Moon, M. A. (2004). *Sales Forecasting Management: A Demand Management Approach*. Thousand Oaks: SAGE Publications.

Park, K. W. K., Lee, S., & Miller, P. (2012, October). Ticket Pricing per Team: The Case of Major League Baseball (MLB). *Allied Academies International Conference: Proceedings of the Ac*, *15(2)*, 21.

Simmons, R. (1996). The demand for English League Football: a club-level analysis. *Applied Economics*¸ *28*, 130-160.

Sloane, P. (1997). Economics of sports: an overview. *Economic Affairs*, *17(3)*, 2-6.

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, Article 4.

Souza, F. A. P. (2004). *Um estudo sobre a demanda por jogos de futebol nos estádios brasileiros.* (Master dissertation). Departamento de Administração, FEA-USP, São Paulo, Brasil.

Szymansky, S. (2001). Income inequality, competitive balance and the attractiveness of team sports: some evidence and a natural experiment from English soccer. *The Economic Journal*, *111*, 69-84.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica, 26*, 67-80.

Van Calster, B. T., Smits, T., & Van Huffel, S. (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)*, 1-22.

Wooldridge, J. M. (2003). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: The MIT Press.