Universidad
de Navarra

Facultad de Ciencias Económicas y Empresariales

Working Paper nº 05/09

Football Betting As A Cyclical Learning Process

Francesc Pujol
ESI, Economics Sport and Intangibles research
Department of Economics
University of Navarra

Football Betting As A Cyclical Learning Process
Francesc Pujol
Working Paper No.05/09
September 2009

ABSTRACT

Based on the well established result of probability matching, we explore in this paper to what extent soccer bettors adjust their behavior taking into account the new relevant information provided by the market. We test empirically the existence of a learning process using the *Quiniela* bettors' performance, the Spanish soccer betting game, for the period 1970-2000. Empirical results suggest that there is a progressive improvement in the bettors' performance attributable to the learning process. The other remaining explanatory variables follow the expected behavior.

Francesc Pujol
ESI
Department of Economics
Universidad de Navarra
Ed. Bibliotecas (Entrada Este)
E-31080 Pamplona
fpujol@unav.es

Football Betting As A Cyclical Learning Process.

Francesc Pujol
ESI, Economics Sport and Intangibles research group
Department of Economics
University of Navarra

ABSTRACT

Based on the well established result of probability matching, we explore in this paper to what extent soccer bettors adjust their behavior taking into account the new relevant information provided by the market. We test empirically the existence of a learning process using the *Quiniela* bettors' performance, the Spanish soccer betting game, for the period 1970-2000. Empirical results suggest that there is a progressive improvement in the bettors' performance attributable to the learning process. The other remaining explanatory variables follow the expected behavior.

keywords: betting market, learning process, soccer, *Quiniela*

JEL classification: D83, L83, D81

Corresponding author:

Francesc Pujol
ESI
Department of Economics
Universidad de Navarra
Ed. Bibliotecas (Entrada Este)
E-31080 Pamplona
Tel: +34948425625.
Fax: +34 948425626.
e-mail: fpujol@unav.es

Football Betting As A Cyclical Learning Process.

1. A Learning Process for Soccer Bettors

Betting games based on sport competitions provide an excellent empirical framework to test a number of economic hypotheses about agents' behavior.

Grant et al (1951), Estes (1957), Arrow (1958) and Siegel and Goldstein (1959) are among the pioneer authors showing the existence of probability matching: the tendency to predict the number of occurrences of a binary event at a rate similar to the actual probability that the event happens. That is, if event A is observed with a probability of 0.6, the individual will choose event A 60 percent of the times. This astonishing result reveals a violation of the rational choice theory if the goal of the game is to guess the correct event. Maximizing rationality should push to choose always event A if the probability of occurrence is higher than 0.5. A substantial amount of literature has studied this phenomenon, always confirming this basic result (see for instance Fantino (1998) or Vulkan (2000) for reviews of the probability matching literature). To some extent, probability matching hypotesis is related with the efficiency of the predicting markets, developped since the internet era (Wolfers and Zitzevitz 2004, 2008). Departures from the probability matching appear when the set of the experiment is modified, allowing for instance for large financial incentives or extensive training (Shanks *et al* 2002).

While probability matching is a puzzle for the rational choice theory, it is paradoxically a condition for an optimal betting market behavior. Matching probability is the best collective choice in sport betting games as far as events with lower probabilities of occurrence are associated to higher betting payoffs, with is the practice followed in all sport betting games. Figlewski (1979) and Asch *et al.* (1982) can be counted among the first authors investigating the effectiveness of the betting market. Much of the literature has focused since then on the analysis of the efficiency of the betting market. Williams (1999) proposes a survey of the literature on this issue. This hypothesis tends to be empirically confirmed in several studies, even if there is always a place for abnormal or non- efficient betting behavior, like the "gambler fallacy" (Terrell 1994).

Departing from the general probability matching rule, it can be observed the tendency to overbet the favorite results, found by several authors, like Woodland and Woodland (2001) for the US National Hockey League, Cain *et al* (2000) for the UK football competition or Osborne (2001) for the US National Football League case. The opposite tends to occur in the horse race betting, where longshots are favored. Golec and Tamarkin (1995) suggest that this behavior is produced by overconfidence, while Hurley and McDonough (1996) estimate that the transaction costs can explain this bias. The favorite overbetting bias should suppose a certain inefficiency of the betting market. Hauser *et al* (2001) attack this conclusion, as they show that the optimal strategy in the soccer betting games is to overbet favorite outcomes as compared to the probability of their occurrence.

All in all, these empirical results tend to suggest that bettors do not move far away from the behavior reflected in the probability matching. It is thus not excessive to suppose that bookmakers' quotas reflect basically the true probability of each possible result of a soccer game. This crucial fact sustains the empirical model that we will propose in this paper in order to investigate the presence of a learning process among bettors.

Recent research efforts have been oriented towards the analysis of the learning process in the betting market. These models are evidently not applied to random games but to gambling games based on sport competitions. In these latter cases, the individuals participating in the betting games have a potential ability progressively to absorb the new information generated by the sport market by its own activity, and should adjust their betting behavior where necessary in order to maximize their chances of success. The theoretical foundations of such learning processes have been studied by Evans and Honkapohja (2001) in a general microeconomic framework, while Laslier *et al* (2001) and Grilli (2001) propose some theoretical foundations based on game theory analysis. Heinemann (2000) propose a learning process model based on neural networks, while Bullard and Duffy (1999) use a genetics algorithm approach.

The learning process has been applied several times to analyze the exchange rates (like Garretsen *et al* 1998), the financial market (see for instance Routledge 1999), or commercial brands (van Osselaer and Janiszewski 2001). Concerning the betting market, it has been applied to the lottery games in Greece (Papachristou and Karamanis 1998) and the UK (Forrest *et al* 2000). See also Gandar et al (1988).

Forrest et al (2005) is an interesting reference paper, as they show that bookmakers' odds predict betters as the season advances. So, statisticians and extremely expert odds setters learn from what happens during the season.

We propose in this paper a further analysis of the learning process by conducting an empirical test for the betting market based on soccer competitions using the *Quiniela* data, the Spanish soccer betting game. Unlike betting markets that hinge on points spreads or odds, the participants in this market merely need to pick outright winners. The key aspect of this market is that participants are not competing against one another, as they would in an odds or point-spread market, so that the betting community as a whole may do poorly or well. This feature of this market makes it an excellent test of learning because one need only to look for more wins over the course of the season, after taking into account the control variables.

Agents introduce new information about the relative power of each team according the results obtained by each team during the season. If bettors are able to collect these additional pieces of information, it should be expected that this increase of knowledge is translated into a better prediction of the results of the matches on average as time goes on during season.

It is important to note that the learning process that we analyze is not continuous, as the interval between two seasons acts as a shock provoking a partial loss of the accumulated knowledge, which raises the level of uncertainty about the accurate betting for each match. This rupture is due to several factors: teams which were identified consistently as weak teams move from *Primera Division* (First Division competition between the best 20 Spanish teams) to *Segunda Division* (Second Division competition with the following 20 best national teams). Simultaneously, the teams identified as top performers in the *Segunda Division* at the end of the season accede to the vacant places left by the outgoing teams. Finally, a considerable number of changes inside each team tend to be concentrated in the interval between two seasons, as teams fire and hire players and coaches. These transformations weaken bettors' knowledge, imposing a readjustment of the assessment of the relative expected sport performance of each team. Of course, not all the gathered knowledge is lost at the beginning of each season, as past sport performance is correlated with future performance.

After having established the chances of each contender, the individual bettor still has to decide in which direction she will place her bet. We have here two profiles: a conservative approach selecting the most probable result or the risky decision supporting the less probable result. If the number of betting agents is sufficiently high and representative, we could expect a similar distribution of votes for each result to the relative chances of appearances, as the empirical results concerning the probability matching robustly suggest.

The main hypothesis of this paper is that if there is a process where information about the market characteristics circulates fairly open and transparently, as it tends to happen in the soccer competition context, then the average percentage of winning bets will tend to increase as the competition advances, *ceteris paribus*. This is because soccer fans accumulate new information helping them to better identify which are the teams that will become the top ranked teams and which will become the bottom liners. The probability of each team winning a given match will be then progressively better estimated and the chances of bettors giving the right result will increase on average as time passes, once controlled for other factors affecting the prediction probabilities.

We will present in the following section the basic facts and mechanisms of the Spanish soccer betting institution. We will empirically test the learning process of soccer bettors in section 3, using the Spanish Quiniela data. Section 4 concludes.

2. The Spanish Soccer *Quiniela*: Some Facts and Figures.

The *Quiniela* was created in Spain as the official public soccer betting institution, organized and managed by a State agency, the *ONLAE (Organizacion Nacional de Loterias y Apuestas del Estado)* in 1946, as a means to generate new complementary public income. The structure and functioning was defined following the example of its Italian sister, the *Tottocalcio*, which had at that time several years of experience and phenomenal popular success.

The game consists of the predictions concerning a set of soccer matches to be played in a given week. The choice for each match is to put a 1 if the forecast is that the team playing at home will win the match; to put an X if a draw is expected and to put a 2 if the prediction is that the team playing away will win the match.

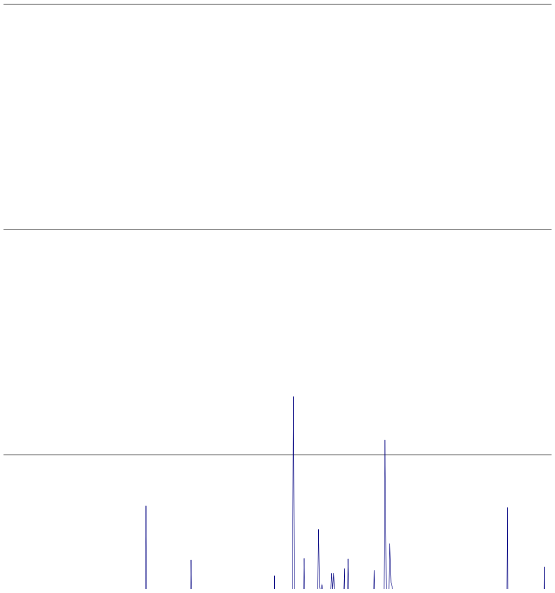
For the very first *Quiniela*, the exact result of seven matches had to be forecast. Soon the number of matches in a week was raised to 14. In 1992 a complementary 15th match to be forecast was introduced, whose result is taken into account for an extra prize only for the bettors who have given the correct prediction for the first 14 matches.

The share of income to be distributed in prices has also followed a very stable system throughout the period. From the total amount of money raised each week, a given percentage is distributed among the people who have predicted correctly the 14 results (in terms of 1, X and 2 choice). Another percentage is distributed for the people having marked correctly all but one result, and the same for the people who have identified correctly 12 results. In 1994 a prize was introduced for the people with just 11 good answers, but this prize is distributed only when the resulting individual prize exceeds a fixed reward.

The cost of each individual slip was fixed at 2 pesetas in 1946, it rose to 3 two years later, to 4 in the year 1961 and other successive increases in prices placed the individual bet at a price of 50 pesetas (some 0.30 euros) in 1998.

The increasing popularity of this game has been constant since the initial 38.530 betting slips played in the birthday of the *Quiniela*. Five years later, there were an average of 1 million bets per week; 10 million since 1963, and more than 100 million bets in the *Quiniela*'s golden age, in the late 70s.

Figure 1

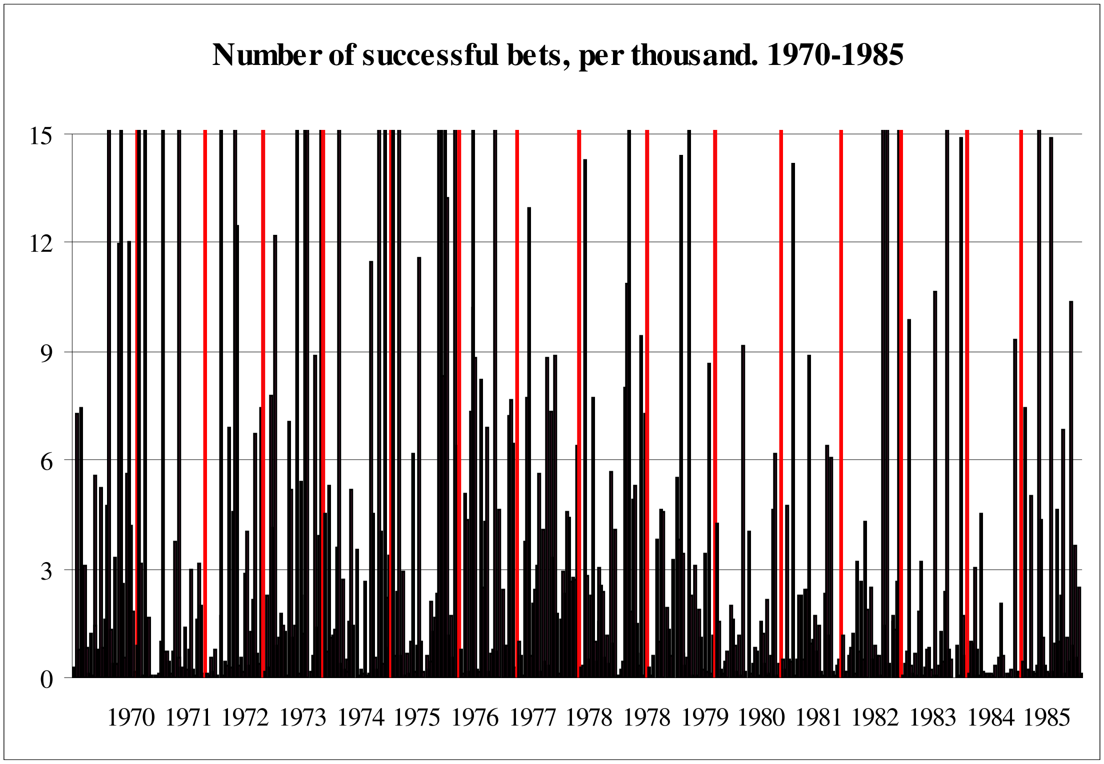


Source: ONLAE, own elaboration

Figure 1 shows the evolution of the number of bets between 1970 and 2000, period which will constitute the basis of our data set for the empirical analysis. At the beginning of the 70s, the average number of bets per week was around 50 million for a total population at that time of some 32 million inhabitants. There is a practically continuous increase in the number of bets, to reach the highest point at the beginning of the eighties, with peaks of more than 150 million bets per week. Between 1982 and 1985 the average is established around one hundred million bets. We observe a dramatic sudden decrease of interest in the *Quiniela*, as the average per week drops to less than 50 millions in 1986, bottoming out around 25 million bets during the late eighties. This phenomenon is mainly due to the introduction of new public betting games not related with sports, at the Central Administration level as well as at the regional level (Autonomous Communities), like the 6/49 game and other similar games based on the selection of certain numbers, which are produced randomly. Forrest et al (2004) observe this phenomenon in betting games in United Kingdom.

The relative resurgence of the popular interest in the *Quiniela* observed during the nineties is linked to the modification of the rules introduced during this period that we mention above. One measure was to give prizes also when indicating correctly 11 results out of 14. The other important measure consists of the introduction of the new 15th complementary match. It produces a double effect: it increases the probability of producing extraordinarily high prizes, as the result is more difficult to attain and, secondly, it creates a reserve fund when there are no bettors who have predicted correctly the 15 results. If this occurs for several consecutive weeks, the increase in the amount for prizes produces a call effect, generating a temporary increase in the number of bets. This phenomenon has been observed with the presence of some isolated peaks since 1992 but was never seen before. The series shows also for the whole of the period some weeks where a pronounced decrease in the amount of betting is observed compared to the average of the period. This is due to the fact that the *Quiniela* is sometimes played even if there are no *Primera Division* matches, for instance, when it is stopped because the national team plays an international match, or when the *Primera Division* matches begin after or end before the matches of the *Segunda Division*. In those cases, the matches concerning the *Primera Division* are replaced by matches of the *Segunda Division* and completed with some matches corresponding to the *Segunda Division B*, an inferior category. Garcia and Rodriguez (2007) show empirically that these variables drive the demand for football pools in Spain.

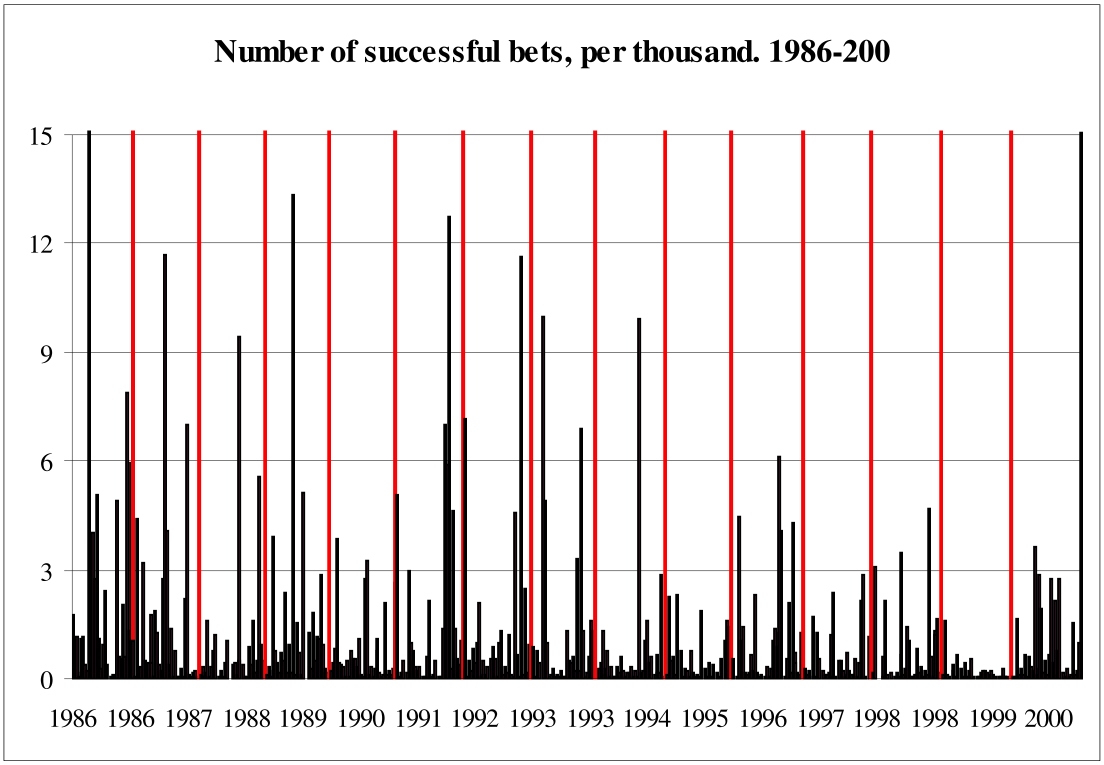
Figure 2



Number of successful bets, per thousand. 1970-1985

<u>Source</u>: ONLAE, own elaboration

Figure 3



Number of successful bets, per thousand. 1986-200

<u>Source</u>: ONLAE, own elaboration

Figures 2 and 3 show the weekly number of successful bets, in per thousand, i.e, the number of slips where at least 12 results were correctly guessed. This is the dependent variable that we will use in the empirical analysis in order to assess the existence of any learning process inside the season. No clear intuitive answer can be given after observing the raw measure of bettors' performance. The division between two seasons is marked by a red line. Depending on the years we can observe an apparent increasing or decreasing path or an apparent absence of relation between weeks inside the season and bettors performance. Then, the learning process, if any, is not apparent and easily evident to external viewers. These figures clearly show that several factors do affect the ability to guess correctly the matches each week, as we will consider in the following section.

We can also appreciate from figures 2 and 3 a long term trend of decreasing levels of betting performance. This phenomenon is not linked to a secular diminution of bettors' knowledge, as this hypothesis is really hard to sustain. We will argue and empirically check in the following section that this evolution is due to structural transformations of the Spanish soccer competition.

In any case, figures 2 and 3 tell us that we have to take into account control variables affecting short term and long term bettors' performance if we want to be able to measure the presence of a learning process inside the soccer betting games.

Having presented the essential facts and figures necessary to understand the functioning of the Spanish Quiniela game we can move to the empirical analysis of soccer bettors' behavior.

3. The Empirical Model Specification and Results

Our main aim is to test whether or not we can retrace the presence of a learning process among the soccer supporters stimulated by the accumulation of matches during the season. We propose to establish the link between this internal process and one observable manifestation of the existence of this learning process: the progressive ability of soccer bettors to adjust their bets to the new information available, translated into an increase in the percentage of people correctly guessing at least 12 results of the weekly *Quiniela*.

Thus, the dependent variable we want to use is the percentage of people having correctly guessed 12, 13 or 14 match results of the *Quiniela*. This information, like all the remaining data used in this section, has been obtained from the data base of the managing organism of the *Quiniela*, the *ONLAE* (*Organizacion Nacional de Loterias y Apuestas del Estado*). Our original data set is formed by the 1257 *Quinielas* played between 1970 and 2000. This dependent variable has been shown in figures 2 and 3 in the precedent section.

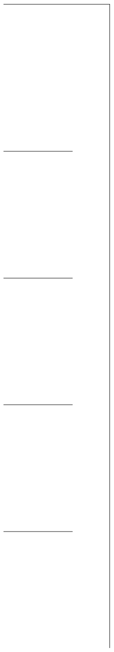
We have proceeded to a first observation cleavage by eliminating the anomalous cases. A first group of deleted observations corresponds to those weeks where one or several of the matches scheduled in the *Quiniela* were not played. In those cases, there are not 14 matches to guess, but fewer.

The second group of eliminated observations, much more important in number, corresponds to those weeks where there are no *Primera Division* matches played. The 14 *Quiniela* results are filled then by *Segunda Division* and *Segunda Division B*, or even with matches of the *Italian*

Scudetto or World Cup matches. It is crucial to drop this kind of *Quiniela*, as they suppose a clear temporary breakdown in the learning process, because the relative power of each team has to be reevaluated for all the new teams occasionally introduced. We have an easy way to identify this kind of *Quiniela* week, as they correspond to a dramatic slowdown of the *Quiniela* number of slips, compared to the previous week. This is because a relevant percentage of people opt not to pay for a *Quiniela* in these weeks, precisely because of bettors' consciousness of the lack of sufficient knowledge to make the forecast with a minimum of chances of having enough correct predictions. All these deletions have reduced the data base to 1063 observations.

Figure 4 shows the distribution of the dependent variable, according to the per thousand *Quiniela* betting slips which have correctly guessed at least 12 out of the 14 matches. The median value is 0.483, that is, some 5 betting slips out of 10 000. In 51.3% of the cases, the per thousand number of winner tickets ranges between 0.08 and 1.28.

Figure 4



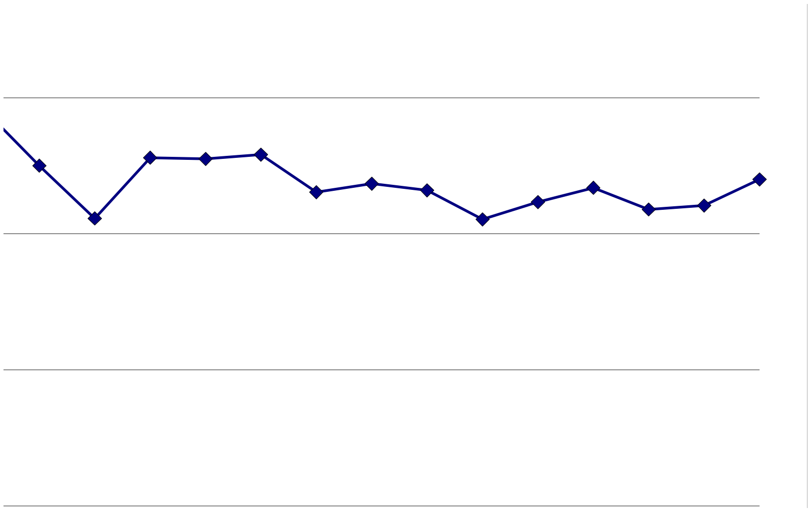
<u>Source</u>: ONLAE, own elaboration

After having characterized the dependent variable, we can move on to the analysis of the explanatory variables we have selected.

In our view, the main factor affecting the chances of generating at least 12 right answers is the aggregate level of difficulty of correctly guessing the 14 match results, which varies significantly from week to week. That is, the most important share of the variance of the performance variable has to be caught by what we can call the intrinsic and the extrinsic difficulty of a given *Quiniela*.

Two main sport factors will help us to understand the source of difficulty of correctly guessing at least 12 matches: the home factor and the rank factor. The first one is the playing-at-home factor (HF). For any given team, independently of its sportive performance, it is easier to win a match played at home than played away, for a number of practical or psychological reasons that we do not need to explore in this paper. The playing-at-home factor can be perceived as a rather stable factor in time and across teams, with a slight variance inside and between seasons, at least for connected seasons. Because of that, it can be expected that this factor is a permanent incorporated knowledge of the characteristics of the market, as it is not lost from one season to another. HF implies that in average *Quinielas* will show more 1 and X than X and 2. We show in figure 5 the extent and the evolution of the HF.

Figure 5



Source: own elaboration

We have named it as “Easiness variable”. It has been built by giving one point for each match won at home in a *Quiniela*, half a point for a draw and zero points for a match won by the visitor team. If the HF was inexistent and the probability of winning a mach was equal in average playing at home or abroad, the “Easiness variable” should take a value of 7 points. Figure 5 shows that the HF has a substantial impact, as it oscillates between 8 and almost 11 points. More specifically, for the period 1994-2001, 63.8% of the points of the Spanish soccer *Liga* have been obtained by teams playing at home, a value clearly bigger than 50% neutral value. We can also observe a progressive weakening of the HF power. The decrease of the HF effect implies a diminution of the probabilities of correctly guessing a match. We can intuitively understand that this structural transformation in the Spanish soccer competition is probably the main factor explaining the evolution of bettors’ performance as shown in figures 2 and 3 in the precedent section.

The second element affecting the intrinsic difficulty of a *Quiniela* is the ranking factor (RF). The ranking factor refers to the difference of sportive performance of the two teams playing a match. By definition, the probability of a top performing team (T) winning a match is higher than the probability of mid performing teams (M) and bottom performing teams (B), and the probability of M winning is higher than B. This almost self-evident assertion is the core and practically only element necessary to model the soccer learning function.

In practical terms, the distance in performance between T and B teams can be estimated by the percentage of points obtained by the top and bottom 4 teams of each year between 1970 and 2000 as shown in figure 6. We do not observe a specific long term trend. Top teams (T) obtain some 60-70% of total possible points, while bottom teams (B) obtain just 40% points.

Figure 6

Source: own elaboration

The correct identification of the RF of each team is the main area where the learning process takes place. An important share of the knowledge concerning the identification of each team RF is lost at the beginning of each season, as explained before, and has to be progressively rebuilt by analyzing the performance of each team during the season. Thus, at the beginning of the season the bettor will count basically with the HF, and will progressively incorporate the information concerning the RF.

The intrinsic difficulty of each *Quiniela* varies strongly every week because the distribution of the difference in RF of the teams playing a match also varies every week, in a random way, as the order of the matches are chosen by a strictly random drawing. A given *Quiniela* can concentrate a high number of matches with teams playing at home against lower RF teams. In this case, RF reinforces HF pushing the winning probabilities of teams playing at home close to one. The overall chances of guessing correctly 12 results increase dramatically. When the opposite occurs, and higher RF teams play the match away in a given week, RF tends to neutralize the HF, moving the probabilities of winning the match at home close to 0.5, the highest level of incertitude. This *Quiniela* will be especially difficult to correctly guess. Summarizing, the higher the number of 1 results (games won at home), the easier will be to predict the *Quiniela* results.

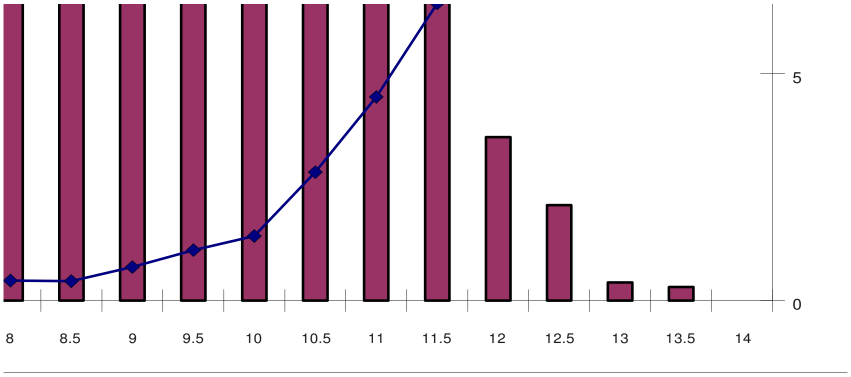
The second element affecting the probabilities for predicting a winning *Quiniela* is the extrinsic difficulty. The extrinsic difficulty of the *Quiniela* is originated by the occurrence of surprising results, that is, the result with lower theoretical probability of occurrence. The extent of this surprise element can also be studied by using the RF analysis. An unexpected match result is one with a theoretical probability of occurrence lower than 0.5. A big surprise corresponds to a result that has a small probability of occurring. The result having the higher probabilities of occurring is a match where a Top team T wins a match played at home against a B Bottom team. When a T team losses this kind of matches, this produces a big surprise, as only a small fraction of bettors' have chosen this unexpected result. If some of these big surprises are concentrated in a given *Quiniela*, the number of winning *Quinielas* will drop dramatically. Here again, a *Quiniela* with high extrinsic difficulty is associated with an accumulation of X and 2 results. The extrinsic factor reinforces difficulty in exactly the same way as the intrinsic difficulty. It is impossible to disentangle the two effects, but this is not a problem for our purpose, as our goal is to capture the difficulty factor of the *Quiniela* in general, in order to isolate the learning process effect that we are looking for.

We could measure the intrinsic difficulty identifying for each one the 1063 *Quinielas* of the sample the standing of each one of the 28 teams (Ranking Factor), thus fixing the expected probability of the 1, X, or 2 result for each match, taking also into account the Home Factor effect. In order to measure the extrinsic difficulty, we should analyze also for each Quiniela the difference between the expected probability and the actual match result.

It is obvious that the effort needed to establish directly the intrinsic and extrinsic difficulty is out of proportion. Fortunately, we have an indirect way to catch this phenomenon. As we have pointed out, both intrinsic and extrinsic levels difficulty are associated with the number of matches won at home, because of the HF and RF. The higher the number of matches won at home (1 result), the easier the Quiniela to be correctly guessed. We can then choose the ex post realization of the results as a proxy for the measure of the ex ante level of intrinsic and extrinsic difficulty. We can propose this association with confidence because of the existence of the probability matching rule: bettors as a whole predict the results in a proportion similar to their actual probability of occurrence.

The empirical results concerning our 1063 observations confirm the approximation we are proposing here, as is shown in figure 7 where we associate the Easiness factor as explained before (where for a *Quiniela* we give a point to a 1 result, 0.5 points to a X result and 0 points to a 2 result), with the chances of having a winning *Quiniela*.

Figure 7



Source: own elaboration

Left scale shows the frequency of each Easiness index value. 9 points, 9.5 and 10 are the most frequent results. The minimum value observed in our series is 4, produced once, while the maximum value is 13.5, a result generated three times. The right scale series shows the per thousand successful Quiniela associated with each Easiness value. A clear positive relation appears between the two variables, as we expected, and the results can be considered representative for the 7 to 11.5 Easiness index values. So, we propose then to measure the intrinsic and extrinsic difficulty of the *Quiniela* betting slips by using the Easiness value of each *Quiniela*.

To measure the robustness of the results, we have introduced a complementary measure to the Easiness index, the Simplified Easiness index, which gives one point to every match won at home, and zero points for all the other results. We prefer the first measure proposed to this one, as it contains the double of different values. Logically, however, both series are intimately linked; the coefficient of correlation of the 1063 observations is 0.91.

The second explanatory variable that we include in the model is the "Week" of the *Liga*, that is the number corresponding to the matches in the season. It takes a value 1 for the first *Quiniela* with *Primera Division* matches of each season; the following *Quiniela* takes the value 2, and so on. This is the key variable designed to catch the presence of a learning process in the soccer betting market. If present, we expect a positive relation between the average performance in the *Quiniela* betting and the accumulation of matches played during the season.

The third independent variable we have included is "Year", that is, the number corresponding to the year when each *Quiniela* is played. We control with this variable the evolution of the role the long-run knowledge factor, mainly determined by the long term evolution of the Home Factor effect. Figure 5 shows that HF has decreased its impact though time. Due to this fact, we expect a negative relationship between the variable “Year” and the dependent variable.

We have also included a couple of control variables. The first one is "Betting Slips", that is, the amount of betting slips paid in a given *Quiniela* week, as shown in the figure 1. We are looking at whether or not the total amount of tickets filled in affects the relative performance of bettors. As we are speaking about a large, mature market, always evolving in terms of tens of millions of bets per week, we assume that these numbers ensure the presence of all kind of bettors, from the most ignorant or uninterested concerning the theoretical probability of each team winning a match, to those who take advantage of all the available information before making their predictions. We expect then the number of betting slips not to have a statistically significant relationship with bettors' performance.

The second control variable is a dummy variable "Three Points", active since 1995, when the new system of points was introduced in the Spanish *Liga*, giving three points for each match won, instead of two points. This change may have produced a structural modification in the system of incentives for the teams playing at home and away. We do not propose an *a priori* guess for the sign, as there are no clear theoretical grounds to define the sense of the relationship.

The basic econometric model is an OLS regression, using the log of the weekly performance as dependent variable, measured as the per thousand bettors having correctly guessed at least 12 results.

Empirical results (column number 1 of the table 1) show that the explanatory power of the proposed model, the adjusted R^2, is about 0.498. As for the explanatory variables we find that, as expected, the variable "Easiness" is strongly positively related with the variable "Performance". The higher the intrinsic and extrinsic Easiness of the Quiniela, measured by the number of matches won or drawn at home, the higher the percentage of people having correctly guessed at least 12 results.

The variable we have selected to count for the existence of a learning process appears also to influence the ability of the bettors to give the right answers. The positive and statistically significant coefficient of the variable "Week" suggests thus that the average performance of the bettors tends to increase progressively during the season, *ceteris paribus*. The higher the number of matches played during the season the higher the mean number of correctly guessed matches per week. We attributed this result to the fact that the advancement of the season enables the soccer fans to better perceive the real power of each team, helping them to better adjust their betting, according to the actual probabilities of each team winning a match. This empirical result tends then to confirm the existence of a learning process in the soccer betting market.

Table 1. Explaining bettors' performance

Dep variable	Perform	Peform	Perform	Perform	Perf net
N. of obs:	1063	1063	1063	1063	1033
Method	OLS	WLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)
C	16.3700	16.3700	17.9775	17.7431	14.1146
	(1.003)	(0.895)	(1.1519	(1.097)	(0.873)
Easiness	0.81697 ***	0.81697 ***		0.81928 ***	0.74957 ***
	(27.429)	(22.063)		(27.755)	(24.845)
Week	0.01210 ***	0.01210 ***	0,00965 ***	0.01355 ***	0.01006 ***
	(3.203)	(3.266)	(2.663)	(3.450)	(2.703)
Year	-0.01350 *	-0.01350	-0.01344 *	-0.01432 *	-0.01276
	(-1.741)	(-1.553)	(-1.912)	(-1.865)	(-1.569)
log(N. Betting Slips)	0.10135 (1.131)	0.10135 (1.131)	0.10019 (1.168)	0.11272 (1.269)	0.10552 (1.197)
Three Points	-0.00264	-0.00264	-0.04892	0.005899	-0.04632
	(-0.018)	(-0.017)	(-0.358)	(0.042)	(-0.328)
1 Result			0.74305 ***		
			(25.258)		
X Result			0.12933 ***		
			(4.108)		
Week 10				0.55242 **	
				(2.243)	
Week 22				-0.66795 ***	
				(-2.737)	
Week 37				-0.66784 **	
				(-2.082)	
Week 40				1.44988 **	
				(2.177)	
Week 42				1.84563 **	
				(2.406)	
Adj R^2	0.4981		0.5409	0.5065	0.4520
F-Stat	209.756 ***		207.381 ***	110.006 ***	171.231 ***
D-W Stat	2.017	2.017	2.045	2.020	1.971

Note: t-Statistic in brackets. *** denotes an interval of confidence of 1%; ** for an interval of confidence of 5%; * for an interval of confidence of 10%.

The remaining independent variables also follow the expected behavior. The variable "Year" is negatively related to the explained variable, in a weak but still statistically significant manner. This confirms the picture given by figure 5 and suggests that there is a developing structural change during the time-period under analysis tending to reinforce the intrinsic difficulty of the *Quiniela*.

As for the control variables, we observe that, as theoretically predicted, the variable "Number of Betting Slips" is not statistically linked with the variable “Performance”. The number of betting slips filled in each week does not affect the relative chances of the bettors obtaining the right results, as there is always a sufficiently high number of bettors capturing all the available information from the soccer market. Finally, the variable "Three Points" is not related to the dependent variable. This result suggests that the change in the rule concerning the points given to each match won has not produced a structural change.

The Durbin-Watson test indicates that there is no serial correlation of the residuals to fear. This is consistent with the nature of our time series, as we do not have any argument to expect that the level of the betting performance one week will influence the betting performance of the following week.

We have estimated some complementary regressions in order to check the robustness of the results. Our first control has been to apply a Weighted Least Square technique using the Newey-West and the White tests, instead of a OLS, in order to inoculate the eventual presence of heteroskedasticity of the errors. We show in column number 2 of table 1 the results concerning the Newey-West test. Results are unchanged concerning the impact of the "Easiness" variable and the "Week" variable, while the coefficient of the "Year" is no longer significant (while it remains narrowly significant with the White test).

As we mentioned previously, we can imagine another reformulation of the variable "Easiness", by splitting the original one in two, the first one, the new variable "1 Result" counting one point for each match won at home in a given *Quiniela*, and a second one, called "X Result", which gives one point for each draw. Column 3 in table 1 contains the coefficients of the new regression, using an OLS model. The results are basically the same: the two new variables play a similar role to the precedent "Easiness" variable, as they are strongly related to the level of performance of the bettors. The variable "Week" maintains its characteristics, influencing positively the level of performance. The remaining variables behave exactly as in the previous specifications.

A third way to test the robustness of the findings consists of a specific treatment of the difficulty factor. We have argued that it is caught by the "Easiness" variable, as we consider that there is a direct relation between the extrinsic and intrinsic Easiness and the number of 1 and X results. Even if this happens on average, both variables do not fit perfectly, as the same number of matches won at home corresponds in some cases to an intrinsic easy *Quiniela* to predict, furthermore with no surprises, while in other cases this same apparent result corresponds to a higher intrinsic difficulty, coupled with more surprising results. As the distribution of matches are organized by blind drawing and as the generation of *Quiniela* surprises depends on the non-realization of very easy-to-predict match results, we suppose that there is a random distribution of the intrinsic and extrinsic difficulty on time. In general we expect that easy and difficult *Quinielas* are equally distributed in time, and we reduce the impact of any given *Quiniela* by the fact that each value of the variable "Week" depends on some 30 observations corresponding to the time period covered. Nevertheless, we cannot exclude finding in practice an abnormal accumulation of specially easy or difficult *Quinielas* in a given week, precisely because the distribution is basically random.

We propose two ways of measuring better the impact of the Easiness influence in order to reinforce further the robustness of the specific role of the "Week" variable, which we presume reflects the learning process influence. The first one consists of the introduction of a dummy variable for each competition week. In this way, we check for the accumulation of especially difficult or especially easy *Quinielas* in a given week. As we look for the weeks having a special performance behavior that cannot be explained by "Easiness" and the other independent variables, we have introduced firstly all the 40 dummy week variables (we have excluded 2 of them with an average value, to avoid the formation of a singular matrix). We have then eliminated one dummy each time from the initial regression with 40 dummies. The dropped dummy week corresponds to the one with the lowest t-Statistic, as this suggests that it does not add an explanatory power to the model and does not follow a specific behavior. The regression in column 4 of table 1 shows the result of the process of dummy elimination at the point where all the remaining weeks are statistically related with the variable Performance, within an interval of confidence of 5%. This affects five weeks. Two of them, number 22 and 37, seem to have attracted more difficulty in predicting the *Quiniela*, while weeks number 10, 40 and 42 concentrate an accumulation of extraordinarily easy *Quiniela*. If we pursue the exercise of elimination of weeks to concentrate on those showing a specific path at the level of confidence of 1%, just the weeks number 22 and 42 are maintained. Concerning the other explanatory variables, none of them suffers any qualitative change because of the introduction of these checks. Furthermore, the impact and the statistical relationship of the variable "Week" with "Performance" is strengthened. The same happened in the different stages of deletion of the redundant week dummies.

The final check for robustness controlling for the impact of the variable Easiness consists of the elimination of the extreme observations in terms of the variable performance. This process pursues the same goal as before, by the isolation of the extreme values of the Easiness factor that could distort the whole picture. If in the preceding strategy we decided to check for the special weeks, here we simply opt to evacuate the problem by dropping the more problematic observations. We have thus eliminated the 1% top and bottom observations. This reduces the range of variation of the dependent variable from the original [0.0009; 166.09] to the new [0.0131; 20.503] interval. We have rerun the initial explanatory model shown in column 1, but with the new data set reduced to 1033 observations. The coefficients are shown in column 5 of table 1. Again, all the original basic results remain unchanged, as the variable "Easiness" is still the most important explanatory variable, and the variable "Week" also helps to understand the level of performance of the *Quiniela* bettors. We have run the same family of complementary regressions as made in the initial data set. There are no relevant changes to mention concerning the explanatory variables.

5. Concluding Remarks

We have followed a long path beginning from the perspective of the existence of a potential learning process in the soccer betting market to arrive at an empirical selection of variables enabling us to test this hypothesis in the case of the Spanish soccer *Quiniela*. We feel that none of the assumptions used in our research is definitively too strong, as some of them have been confirmed in many other experimental and empirical studies. We think thus that the empirical results can be viewed as a fair approximation to the issues studied.

These results suggest that there is in fact a process of integration of new information concerning the different strength of soccer teams, which is translated into an increase in the level of performance of soccer bettors to correctly guess at least 12 out of 14 match results, as the soccer season evolves, *ceteris paribus*. The presence of a learning process is far from evident when looking directly at the raw measure of bettors' performance (figures 2 and 3). Its statistical relevant presence appears only when taking into account all other variables influencing bettors' performance. It is worth to note the fact that a market functioning basically inside the framework of effectiveness due to the presence of probability matching can still increase its internal performance by incorporating new information generated during the season thanks to the learning process.

The robustness of the presence of the learning process in this soccer betting game suggest that learning processes should be considered as a relevant factor to consider in any market where consumers' decision can be influenced by the production of new information.

Finally, the presence of a learning process, as well as the coherent behavior of the remaining explanatory variables, tends to suggest that it is possible to understand to some extent the functioning of the soccer betting market. This secondary result of the paper opens the way to new related research using the Spanish *Quiniela* and soccer betting games in general. Astonishing, this kind of studies are explored to a lesser extent than the research focused on the analysis of other random betting games. The analysis of the agents' behavior and predictions based on non-random games potentially generates some economic and psychological knowledge that cannot be attained by analyzing random games.

References

Arrow, K (1958). Utilities, Attitudes, Choices: A Review Note, *Econometrica*, 26, pp. 1-23.

Asch, P., Malkiel, B. and Quandt, R., (1982). Racetrack Betting and Informed Behavior. *Journal of Financial Economics*, 10(2); pp. 187-194.

Blume, L. and Easley, D. (1982). Learning to be Rational. *Journal of Economic Theory* 26, pp. 340-351.

Bullard, J. and Duffy J., (1999). Using Genetics Algorithms to Model the Evolution of Heterogeneous Beliefs. *Computational Economics* 13(1), pp. 41-60.

Bray, M., (1982). Learning, Estimation, and Stability of Rational Expectations. *Journal of Economic Theory*, 26, pp. 318-339.

Cain, M., Law, D. and Peel, D., (2000). The Favourite Longshot Bias and Market Efficiency in UK Football Betting. *Scottish Journal of Political Economy*, 47(1), pp. 25-36.

Estes W. (1957), "*Theory of learning with constant, variable or contingent probabilities of reinforcement*",Psychometrika, 22.

Evans, G. and Honkapohja, S., (2001). *Learning and Expectations in Macroeconomics*, Princeton University Press, Princeton.

Fantino, E. (1998). Behavioral Analysis and Decision Making. *Journal of Experimental Analisys of Behavior*, 69, pp. 355-364.

Figlewski, S., (1979). Subjective Information and Market Efficiency in a Betting Market. *Journal of Political Economy*, 87(1), pp. 75-88.

Forrest, D., Gulley, O. and Simmons, R., (2000). Testing for Rational Expectations in the UK National Lottery. *Applied Economics*, 32(3), pp. 315-326.

Forrest, D., Gulley, O. and Simmons, R., (2004), Substitution Between Games in the UK National Lottery, *Applied Economics*, 36, pp. 645-651.

Gandar, J., W. Dare, C. Brown and R. Zuber, (1988). Informed Traders and Price Variations in the Betting Market for Professional Basketball Games. *Journal of Finance*, 53(1), 385-401.

Garcia, J. and Rodriguez, P. (2007). The Demand for Football Pools in Spain, *Journal of Sports Economics*, 8(4), pp. 335-354.

Garretsen, H., Knot, K. and Nijsse, E., (1998). Learning about Fundamentals: The Widening of the French ERM Bands in 1993. *Weltwirtschaftliches Archiv/Review of World Economics*, 134(1), pp. 25-41.

Golec, J. and Tamarkin, M., (1995). Do Bettors Prefer Long Shots because They are Risk-Lovers, or Are They Just Overconfident?. *Journal of Risk and Uncertainty*, 11(1), pp. 51-64.

Grant D.A., Hake H.W., and Hornseth J.P. (1951), "*Acquisition and extinction of verbal expectations in situation analogous to conditioning*", Journal of Experimental Psychology, 42, 1-5.

Grilli, M., (2001). A General Approach to Rational Learning in Games. *Bulletin of Economic Research*, 53(4), pp. 275-303.

Hauser, U., Konig, U. and Krylova, E., (2000). Gamblers' Rationality in Parimutuel Soccer Betting. *Economics Series, Institute for Advanced Studies (IHS),* N. 87.

Heinemann, M., (2000). Adaptative Learning of Rational Expectations Using Neural Networks. *Journal of Economic Dynamics and Control*, 24(5-7), pp. 1007-1026.

Hurley, W. and Mcdonough, L., (1996). The Favourite Longshot Bias in Parimutuel Betting: A Clarification of the Explanation That Bettors Like to Bet Longshots. *Economics Letters*, 52(3), pp. 275-78.

Laslier, J.-F., Topol, R. and B. Walliser, B., (2001). A Behavioral Learning Process in Games. *Games and Economic Behavior*, 37(2), pp. 340-366.

Osborne, E., (2001). Efficient Markets? Don't Bet on It. *Journal of Sports Economics*, 2(1), pp. 50-61.

Papachristou, G. and Karamanis, D., (1998). Investigating Efficiency in Betting Markets: Evidence from the Greek 6/49 Lotto. *Journal of Banking and Finance*, 22(12), pp. 1597-1615.

Routledge, B., (1999). Adaptative Learning in Financial Markets. *Review of Financial Studies*, 12(5), pp. 1165-1202.

Siegel, S. and D. Goldstein (1959). Decision-Making Behavior in a Two-Choice Uncertain Outcome Situation. *Journal of Experimental Psychology*, 57, pp. 37-42.

Shanks, D. R. Tunney and J. McCarthy (2002). A Re-Examination of Probability Matching and Rational Choice. *Journal of Behavioral Decision Making*, 15, pp. 233-250.

Terrell, D., (1994). A Test of the Gambler's Fallacy: Evidence from Pari-mutuel Games. *Journal of Risk and Uncertainty*, 8(3), pp. 309-317.

van OSSELAER, S. and JANISZEWSKI, C., (2001). Two Ways of Learning Brand Associations. *Journal of Consumer Research*, 28(2), pp. 202-203.

Vulkan, N. (2000), An Economist's Perspective on Probability Matching*, Journal of Economic Surveys*, 14, pp. 101-118.

Williams, L., (1999). Information Efficiency in Betting Markets: A Survey. *Bulletin of Economic Research*, 51(1), pp. 1-30.

Wolfers, J. and Zitzevitz, E. (2004), Prediction Markets, *Journal of Economic Perspectives*, 18(2), pp. 107-126.

Wolfers, J. and Zitzevitz, E. (2008), Prediction Markets in Theory and Practice, in Durlauf, S. and Blume, E. (eds.), *The New Palmgrave Dictionary of Economics*, Palmgrave Macmillan.

Woodland, L. and Woodland, B., (2001). Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots?. *Southern Economic Journal*, 67(4), pp. 983-995.